

Weismayer, Christian

Book — Digitized Version

Statische und longitudinale Zufriedenheitsmessung

Forschungsergebnisse der Wirtschaftsuniversität Wien, No. 48

Provided in Cooperation with:

Peter Lang International Academic Publishers

Suggested Citation: Weismayer, Christian (2011) : Statische und longitudinale Zufriedenheitsmessung, Forschungsergebnisse der Wirtschaftsuniversität Wien, No. 48, ISBN 978-3-631-75396-5, Peter Lang International Academic Publishers, Berlin, <https://doi.org/10.3726/b13916>

This Version is available at:

<https://hdl.handle.net/10419/182878>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

FORSCHUNGSERGEBNISSE DER
WIRTSCHAFTSUNIVERSITÄT WIEN

CHRISTIAN WEISMAYER

Statische und longitudinale Zufriedenheitsmessung



PETER LANG

CHRISTIAN WEISMAYER

Statische und longitudinale Zufriedenheitsmessung

Dieses Buch beschreibt eine Longitudinalstudie aus dem Bereich der Zufriedenheitsmessung. Auf Basis von Zufriedenheitsdaten bezüglich einzelner Serviceleistungsattribute von Kinos werden Eigenschaften, Auswirkung und Veränderung von Zufriedenheits- und Wichtigkeitsbewertungen gemessen. Ausgehend von der literaturbasierten Entwicklung unterschiedlicher Messmethodik im Zeitablauf, diente die Panelstudie vorwiegend dazu, Veränderungen von Zufriedenheitsmerkmalen zu messen. Ein buntes Sortiment unterschiedlicher quantitativer Herangehensweisen wie Paarvergleiche, multiple Korrespondenzanalyse, Rasch-Modelle, Mehrebenenmodelle, latente Wachstumskurvenmodelle, oder Markov-Kettenmodelle, zeigt longitudinale und vereinzelt auch statische Eigenschaften der Teilzufriedenheitsbewertungen auf.

Christian Weismayer studierte zwischen 1999 und 2010 Betriebswirtschaft an der Wirtschaftsuniversität Wien und arbeitete am Institut für Tourismus und Freizeitwirtschaft in den Jahren 2006 bis 2011.

Statische und longitudinale Zufriedenheitsmessung

Forschungsergebnisse der Wirtschaftsuniversität Wien



Band 48



PETER LANG

Frankfurt am Main · Berlin · Bern · Bruxelles · New York · Oxford · Wien

CHRISTIAN WEISMAYER

Statische und longitudinale Zufriedenheitsmessung



PETER LANG

Internationaler Verlag der Wissenschaften

Bibliografische Information der Deutschen Nationalbibliothek

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

Open Access: The online version of this publication is published on www.peterlang.com and www.econstor.eu under the international Creative Commons License CC-BY 4.0. Learn more on how you can use and share this work: <http://creativecommons.org/licenses/by/4.0>.



This book is available Open Access thanks to the kind support of ZBW – Leibniz-Informationszentrum Wirtschaft.

Gefördert durch die **Wirtschaftsuniversität Wien**.

Umschlaggestaltung:
Atelier Platen, nach einem Entwurf
von Werner Weißhappl.

Universitätslogo der **Wirtschaftsuniversität Wien**:
Abdruck mit freundlicher Genehmigung
der **Wirtschaftsuniversität Wien**.

Gedruckt auf alterungsbeständigem,
säurefreiem Papier.

ISSN 1613-3056
ISBN 978-3-631-60765-7
ISBN 978-3-631-75396-5 (eBook)

© Peter Lang GmbH
Internationaler Verlag der Wissenschaften
Frankfurt am Main 2011

Großer Dank gebührt

Herrn Josef Mazanec, für die Möglichkeit in den Jahren 2006-2010 an seinem Institut für Tourismus und Freizeitwirtschaft zu arbeiten, welche ich als lehrreiche und interessante Zeit genossen habe,

Herrn Reinhold Hatzinger für die anregenden Kurse im Bereich psychometrischer Messmethoden,

dem gesamten Projektteam für die Zeit vor und im Zuge der Projektarbeiten

und meiner Familie für die Unterstützung am gesamten Lebensweg.

Alle Segnungen unserer modernen Zivilisation beruhen schließlich auf der Fähigkeit des Menschen, seinen Horizont zu beschränken, und nicht, wie gemeinhin behauptet wird, ihn zu erweitern.

Ja, man kann mit Fug und Recht behaupten, dass der Nobelpreis die höchste Prämierung der dichtesten Scheuklappen ist, eine Ehrung für konsequente, beharrliche Engstirnigkeit.

Stefan Slupetzky

(Lemmings Zorn, Lemmings vierter Fall, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg, April 2009)

Vorwort

Die folgende Arbeit beschreibt Erkenntnisse aus dem Projekt ‚CINEMON‘. Die Vorbereitungsarbeiten unter dem Projekttitel ‚CISAMON‘ wurden durch den Jubiläumsfonds der Stadt Wien finanziert und beschäftigten sich im Halbjahr zwischen 15. März und 15. September 2006 mit der Planung und dem Aufbau der Studie. ‚CINEMON‘ stellt das Nachfolgeprojekt und gleichzeitig den Hauptteil des Kinoforschungsvorhabens finanziert durch die Österreichische Forschungsförderungsgesellschaft mbH (FFG) im Zeitraum zwischen 1. Oktober 2006 und 31. März 2008 dar. Von Oktober 2006 bis September 2007 wurde mittels online-basiertem Fragebogen das Besucherverhalten in den Wiener Kinos erfragt. Im Anschluss folgte eine im Brückenschlagsprogramm der FFG bestimmte Nachbearbeitung der gewonnenen Daten.

Inhaltsverzeichnis

1.	Abkürzungsverzeichnis	17
2.	Forschungsziele	21
3.	Ausgewählte Konzepte aus der Zufriedenheitsliteratur	23
3.1.	Zufriedenheit und Wichtigkeit	23
3.1.1.	Human Motivation	23
3.1.2.	Importance – Performance	24
3.1.3.	Eindimensional vs. zweidimensional	30
3.1.4.	Satisfier - Dissatisfier	31
3.1.5.	Kano-Modell	33
3.1.6.	Asymetrischer Einfluss	37
3.1.7.	Nicht-linearer Zusammenhang	39
3.1.8.	Teilzufriedenheit - Gesamtzufriedenheit	43
3.2.	Modellvergleich	46
3.2.1.	Importance Grid vs. Penalty-Reward-Contrast Analysis	46
3.2.2.	Importance-Performance Matrix vs. Dummy-Regression	52
3.2.3.	Zufriedenheitskonstrukte	54
3.3.	Segmentierung	58
3.3.1.	Spezifizierung eines erweiterten Kano-Modells	58
3.4.	Konstrukte und Dimensionen höherer Ordnung	60
3.4.1.	Wichtigkeitsdimensionen	60
3.4.2.	Zufriedenheitsherleitung	63
3.4.3.	Delight	65
3.4.4.	Kognitive und affektive Beeinflussung der Zufriedenheit	68
3.4.5.	Berücksichtigung der SERVQUAL Faktoren	70
3.4.6.	Eine polynomiale Lösung	70
3.4.7.	Faktoreneinteilung unter zusätzlichen Annahmen	74
3.4.8.	Nichtlinearitäten und Interaktionen	76
3.5.	Statische Effekte	77
3.5.1.	Zufriedenheit und Loyalität	77
3.5.2.	Performance, Zufriedenheit und zukünftiges Verhalten im Zeitablauf	78
3.6.	Longitudinale Analysen	80
3.6.1.	Longitudinale Effekte	80
3.6.2.	Zufriedenheit und Profit	82
3.6.3.	Zufriedenheit, Wichtigkeit und Verhalten	83
3.6.4.	Longitudinale Messung der Servicequalität	84
3.6.5.	Kundenzufriedenheit und Treue	84
3.6.6.	Veränderung der Zufriedenheit	86

3.7.	Empirische Überprüfung der statischen Konzepte	86
4.	Empirische Methodenüberprüfung und -entwicklung	89
4.1.	Deskriptive Statistik	89
4.2.	Literaturbasierte Konzepte	90
4.2.1.	Importance-Performance Analyse	90
4.2.2.	Dummy-Regression	92
4.2.3.	Vavras-Importance-Grid	96
4.2.4.	Zusammenhang zwischen Wichtigkeit und Zufriedenheit	102
4.3.	Neue methodische Herangehensweisen	110
4.3.1.	Paarvergleichsmodelle	110
4.3.1.1.	Statische Paarvergleichsmodelle anhand der Kinodaten	112
4.3.1.2.	Longitudinale Paarvergleichsmodelle - Kinodaten	116
4.3.2.	<i>Item Response Theory: vom Rasch model zum linear logistic model with relaxed assumptions</i>	121
4.3.2.1.	<i>Rasch Model - Kinodaten</i>	145
4.3.2.2.	<i>Modellevaluierung - Kinodaten</i>	148
4.3.2.3.	<i>Rating Scale Model - Kinodaten</i>	151
4.3.2.4.	<i>Partial Credit Model - Kinodaten</i>	153
4.3.2.5.	<i>Gruppenvergleiche - Kinodaten</i>	164
4.3.2.6.	<i>Linear Logistic model with Relaxed Assumptions - Kinodaten</i>	165
4.3.3.	Multiple Korrespondenzanalyse	169
4.3.3.1.	<i>Homals - Kinodaten</i>	174
4.3.4.	<i>Multilevel Modelling: vom fixed effects model zum multilevel growth curve model</i>	186
4.3.4.1.	<i>Random intercept fixed slope model - Kinodaten</i>	191
4.3.4.2.	<i>Random intercept random slope model - Kinodaten</i>	196
4.3.4.3.	<i>Non-linear random intercept random slope model - Kinodaten</i>	199
4.3.4.4.	<i>Externe Parameter und Interaktionseffekte - Kinodaten</i>	204
4.3.4.5.	<i>Varianzvisualisierungen der unterschiedlichen Ebenen - Kinodaten</i>	207
4.3.4.6.	<i>Modellaufbau mit drei Ebenen - Kinodaten</i>	214
4.3.4.7.	<i>Residuenbetrachtungen - Kinodaten</i>	219
4.3.4.8.	<i>Intercept-slope-Zusammenhang - Kinodaten</i>	225
4.3.4.9.	<i>Behandlung ordinaler Skalen - Kinodaten</i>	238
4.3.4.10.	<i>Quadratische und kubische Terme - Kinodaten</i>	238
4.3.4.11.	<i>Ordered multinomial model - Kinodaten</i>	239
4.3.4.12.	<i>Die Zeit als eigene Ebene - Kinodaten</i>	243
4.3.4.13.	<i>Linear growth curve model - Kinodaten</i>	244

4.3.4.14.	Nicht-Linearitäten - Kinodaten	246
4.3.4.15.	Übergang von der regressionsbasierten zur kovarianzbasierten Lösung	252
4.3.5.	Faktorenermittlung - Kinodaten	257
4.3.6.	<i>Growth Curve Modelling</i> : vom <i>latent growth curve model</i> zum <i>finite growth mixture model</i>	259
4.3.6.1.	Modellspezifikation	265
4.3.6.2.	Schätzung von Modellen mit <i>missing values</i>	274
4.3.6.3.	<i>Latent growth curve model</i> mit linearem <i>slope</i> - Kinodaten	277
4.3.6.4.	<i>Latent growth curve model</i> mit quadratischem <i>slope</i> - Kinodaten	282
4.3.6.5.	<i>Latent growth curve model</i> mit frei geschätzten <i>slope</i> -Parametern - Kinodaten	282
4.3.6.6.	<i>Piecewise latent growth curve model</i> - Kinodaten	284
4.3.6.7.	<i>Parallel latent growth curve model</i> mit Effekten - Kinodaten	286
4.3.6.8.	Kinostandortspezifische Veränderungen und Fehlerwertverzerrungen - Kinodaten	287
4.3.6.9.	Lineare vs. nicht-lineare Wachstumsverläufe - Kinodaten	293
4.3.6.10.	Indikatorspezifisch zeitübergreifende diachrone Korrelationen - Kinodaten	295
4.3.6.11.	Modellfitveränderungen - Kinodaten	296
4.3.6.12.	Einführung von Zeitpunktparametern - Kinodaten	298
4.3.6.13.	Zusammenhang zwischen <i>intercept</i> und <i>slope</i> - Kinodaten	299
4.3.6.14.	A-posteriori ermittelte Populationsunterschiede	301
4.3.6.15.	Fehlerwertbehandlung	314
4.3.6.16.	<i>Finite growth mixture model</i> - Kinodatensatz	316
4.3.7.	<i>Markov Chain Modelling</i> : von der <i>latent class analysis</i> zum <i>mixed latent Markov chain model</i>	318
4.3.7.1.	Anwendungsbeispiele und Spezifizierungen	320
4.3.7.2.	Die Modellierungsumgebung - <i>General Growth Mixture Modelling</i>	324
4.3.7.3.	Modellierungsmöglichkeiten	326
4.3.7.4.	Klassische Modellhierarchie vom <i>mixture latent Markov model</i> ausgehend	329
4.3.7.5.	Modellspezifikationen des umfassenden <i>mixture latent Markov model</i>	330
4.3.7.6.	Beispiele der <i>latent transition analysis</i>	331
4.3.7.7.	Die Messmodellebene	334

4.3.7.8.	Populationsunterschiede	335
4.3.7.9.	<i>Nonstationary manifest Markov model</i> - Kinodatenatz	336
4.3.7.10.	<i>Stationary manifest vs. stationary latent Markov model</i> - Kinodatenatz	338
4.3.7.11.	<i>Cross-sectional latent class models</i> - Kinodatenatz	338
4.3.7.12.	<i>Latent (hidden) Markov model</i> - Kinodatenatz	340
4.3.7.13.	<i>Latent transition analysis</i> – Beispiele aus der Literatur	342
4.3.7.14.	<i>Transition probability matrix</i> - Spezifizierung	346
4.3.7.15.	Latente Ebenen	347
4.3.7.16.	Modellfitvergleiche der <i>latent class analysis</i> - Kinodatenatz	349
4.3.7.17.	<i>Partial measurement invariance vs. full measurement invariance vs. full measurement noninvariance</i>	352
4.3.7.18.	<i>First-order vs. Autokorrelationen höherer Ordnung (second-order, third-order, ...)</i>	353
4.3.7.19.	Modellvergleichsmöglichkeiten und Einschränkungen der <i>transition matrix</i>	354
4.3.7.20.	Modellvergleich unter Verwendung unterschiedlicher Zeitspannen - Kinodatenatz	355
4.3.7.21.	<i>Stationary vs. nonstationary latent transition analysis</i> - Kinodatenatz	357
4.3.7.22.	Modellvergleich mit unterschiedlicher Klassenanzahl im Messmodell - Kinodatenatz	357
4.3.7.23.	<i>Non-stationary two-class model vs. stationary three-class model</i> - Kinodatenatz	358
4.3.7.24.	<i>Non-stationary two-class model mit full measurement invariance</i> - Kinodatenatz	359
4.3.7.25.	Kovariaterweiterungen	362
4.3.7.26.	<i>Longitudinal Guttman Simplex</i>	368
4.3.7.27.	<i>Mixed Markov model</i>	372
4.3.7.28.	Modellselektion	374
4.3.7.29.	Übergänge zwischen unterschiedlichen Populationen	375
4.3.7.30.	<i>Mixed Markov latent class model</i>	376
4.3.7.31.	<i>Mover vs. stayer</i>	377
4.3.7.32.	<i>Latent Markov model</i> - Spezifikation	379
4.3.7.33.	Modelleigenheiten und Kovariate	382
4.3.7.34.	Modellhierarchie	383
4.3.7.35.	Interpretationshinweise - Kinodatenatz	385
4.3.7.36.	Verschachtelung und Modellidentifikation	387
4.3.7.37.	Modellvergleich - Kinodatenatz	388

4.3.7.38.	Modellerweiterungen	389
4.3.7.39.	Die latente Kovariate in Form einer latenten Klassenvariable	391
4.3.7.40.	<i>Mover-stayer model</i> - Kinodatensatz	393
5.	Schlussfolgerungen und Verwertbarkeit	397
6.	Diagrammverzeichnis	401
7.	Tabellenverzeichnis	406
8.	Formelverzeichnis	410
9.	Modellverzeichnis	411
10.	Literaturverzeichnis	413

1. Abkürzungsverzeichnis

2PL	Two-Parameter Logistic
AIC	Akaike Information Criterion
ALS	Alternating Least Squares
AMIR	modifiziertes Missing Information Ratio
A-OSL	Attributes-Overall Satisfaction Link
ARC	Affective Response to Consumption scale
BaF	Basisfaktoren
BeF	Begeisterungsfaktoren
BIC	Bayes Information Criterion
BLRT	Bootstrap Likelihood Ratio Test
CAIC	Consistent Akaike Information Criterion
CBRF	Category Boundary Response Function
CCC	Category Characteristic Curve
CD-Methode	Confirmation/Disconfirmation-Methode
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CIT	Critical Incident Technique
CLR	Conditional Likelihood Ratio
cML	conditional Maximum Likelihood
CS	Customer Satisfaction
CTT	Classical Test Theory
D-T scale	Delighted-Terrible scale
DSM	Diagnostic and Statistical Manual of mental disorders
eRm	extended Rasch modelling
EBTM	Extended Bradley-Terry-Modell
FIML	Full Information Maximum Likelihood
FMA	Factor Mixture Analysis
FMM	Finite Mixture Model
FQL	Full Quasi Likelihood
GGMM	General Growth Mixture Model
GHLM	Generalized Hierarchical Linear Models
GLR	Generalized Likelihood Ratio
GMM	Growth Mixture Model
GRM	Graded Response Model
HLLRA	Hybrid Linear Logistic model with Relaxed Assumptions
I	Intercept
ICC	Item Characteristic Curve
ICOMP	Informational Complexity Criterion

ICRF	Item Category Response Function
IGLS	Iterative Generalized Least Squares
IPA	Importance Performance Analyse
IRF	Item Response Function
IRT	Item Response Theory
LCA	Latent Class Analysis,
LCFA	Latent Class Factor Analysis
LCGA	Latent Class Growth Analysis
LCGM	Latent Class Growth Models
LF	Leistungsfaktoren
LGEE	Logistic Generalized Estimation Equations
LGM	Latent Growth Model
LGMM	Latent Growth Mixture Model
LGS	Longitudinal Guttman Scale
LLRA	Linear Logistic model with Relaxed Assumptions
LLTM	Linear Logistic Test Model
LM	Latent Markov
LMM	Latent Mixed Markov
LMR-LRT	Lo, Mendell and Rubin Likelihood Ratio Test
LPCM	Linear Partial Credit Model
LQL	Listwise present deletion Quasi Likelihood
LRSM	Linear Rating Scale Model
LRT	Likelihood Ratio Test
LTA	Latent Transition Analysis
MAR	Missing At Random
MCAR	Missing Completely At Random
MCMC	Markov Chain Monte Carlo
M-GRM	Modified Graded Response Model
MIR	Minimum Information Ratio
ML	Maximum Likelihood
MLM	Mixture Latent Markov (im Mehrebenkapitel)
MLM	MultiLevel Models (im Markovkapitel)
MLPM	Mixed Latent Partial Markov chain model
MLR	MultiLevel Regression
MM	Mixed Markov
mML	marginal Maximum Likelihood
MMLC	Mixed Markov Latent Class
MRM	Mixed Rasch Modell
MRMLC	Multidimensional Rasch Model for Learning and Change
NI	NonIgnorable

NNFI	NonNormed Fit Index
OLS	Ordinary Least Square
P	Probability
PCA	Principal Component Analysis
PCM	Partial Credit Model
PIF	Performance Importance Function
PPA	Pairwise Present Approach
PQL	Penalized Quasi Likelihood
Pr	Probability
PRC	Penalty Reward Contrast
Quasi-ML	Quasi-Maximum Likelihood
RD	Reading Disability
REML	Restricted Maximum Likelihood
REML	Restricted iterative generalized least squares
RIGLS	Restricted Iterative Generalized Least Squares
RML	Restricted Maximum Likelihood
RMSEA	Root Mean Square Error of Approximation
RMSR	Root Mean Squared Residual
RSM	Rating Scale Model
S	Slope
SD	Standard Deviation
SEM	Structural Equation Model
Sig.	Signifikanzniveau
SBC	Schwartz's Bayesian Criterion
SLC	Structured Latent Curve
SOM	Self-Organizing Map
SPGB	Semi Parametric Group Based
SRMR	Standardized Root Mean square Residual
SSABIC	Sample-Size Adjusted Bayes Information Criterion
TCOS	Kundengesamtzufriedenheit
TD	Typical Development
TLI	Tucker Lee Index
TOCA-R	Teacher Observation of Classroom Adaptation - Revised
TRN	Topology Representing Network
uML	unconditional Maximum Likelihood
VIF	Varianzinflationsfaktor
VPC	Variance Partition Coefficient
WLS	Weighted Least Squares
WLSMV	Weighted Least Square estimator with Variance and Mean adjusted chi-square

WRMR
ZB

Weighted Root Mean square Residual
Zufriedenheitsbeitrag

2. Forschungsziele

Vorliegendes Werk beschreibt die Analyseschwerpunkte und Ergebnisse der Longitudinalstudie. Auf Basis von Zufriedenheitsdaten bezüglich einzelner Serviceleistungsattribute der Kinos werden Eigenschaften von Zufriedenheitsbewertungen gemessen, Auswirkung und Veränderung dieser, und die Wichtigkeit der betrachteten Teilzufriedenheitsbewertungen. Ausgehend von der literaturbasierten Entwicklung und Veränderung unterschiedlicher Messmethoden im Zeitablauf, werden Eigenschaften von Zufriedenheitsbewertungen aufgedeckt. Unterschiedliche Serviceleistungsattribute und deren Auswirkung auf die Gesamtzufriedenheit werden deren Wirkungsweise verdeutlichen. Die Panelstudie diente dazu, Veränderungen von Zufriedenheitsmerkmalen zu messen und beschreibbar zu machen. Diese Veränderungen werden im Zuge der Analysen ebenfalls erklärbar gemacht.

Panelstudien sind in der Zufriedenheitsforschung rar. Die zeitabhängigen Gedanken des Kano-Modells zu untersuchen, wurde zwar oft appelliert, aber bis dato nicht ausreichend durchgeführt. Diesem Umstand wird hier Abhilfe durch variantenreiche Methoden geschaffen. Die auf den erzielten Ergebnissen aufbauenden praxisrelevanten Interpretationen und Überlegungen bleiben größtenteils individuell jedem einzelnen Leser überlassen. Die Arbeit dient dazu, konzeptionell begründete Einblicke in die Welt der Zufriedenheitsforschung zu bekommen und das Blickfeld für methodisch tiefergehende longitudinale Überlegungen zu öffnen und zu formen. Das Hauptaugenmerk liegt folglich in der Vorstellung variantenreicher statistischer Methoden.

Im Wesentlichen ist das Werk in drei Hauptteile untergliedert. Kapitel 2 bespricht die theoretischen Hintergründe der Forschungsaspekte, Kapitel 3 stellt den empirischen Teil dar. Letzterer ist in zwei Unterkapitel unterteilt, nämlich Kapitel 3.2, welches zur Überprüfung der bereits in der Literatur aufgefundenen Methoden dient, und Kapitel 3.3, welches in neuartige Methoden einführt, die bis dato in der Zufriedenheitsforschung nicht verwendet wurden.

Um einen besseren Überblick zu schaffen wird hier ein detaillierter Überblick zum Inhaltsverzeichnis angeführt. Kapitel 2 beschreibt die relevantesten in der Literatur auffindbaren Artikel und ist auf Basis inhaltlicher Strömungen aufgebaut. Kapitel 2.1 bietet eine allgemeine Einführung in den Bereich der Zufriedenheitsforschung und dessen Zusammenhang zur Wichtigkeit. So werden die relevantesten Methoden und die darin aufgegriffenen Aspekte besprochen und im Kapitel 2.2 schließlich einige davon miteinander verglichen. Kapitel 2.3

weist auf die Notwendigkeit der Segmentierung hin, Kapitel 2.4 auf die messtechnische Verbesserung aufgrund der Verwendung latenter Konstrukte. Kapitel 2.5 und 2.6 beschreiben abschließend kurz die statischen und eher spärlich untersuchten longitudinalen Effekte der Zufriedenheit. Kapitel 3 ist in drei Hauptteile untergliedert. Kapitel 3.1 liefert einen kurzen deskriptiven Einblick in den Datensatz. Allgemeine Aussagen dazu aus späteren Analysen müssen jedoch mit Vorsicht betrachtet werden. Gründe dafür werden in den darauf folgenden Kapiteln an den jeweils relevanten Stellen angemerkt. In Kapitel 3.2 werden einige Methoden durchgespielt, welche in der Literatur aufgefunden wurden. Darauf folgen im gesamten Kapitel 3.3 Vorschläge des Autors zur Zufriedenheitsanalyse. Paarvergleichsmodelle, um auch die Wichtigkeit über die Zeit hinweg zu untersuchen, bilden im Kapitel 3.3.1 den Anfang. Darauf folgen die Kapitel 3.3.2 und 3.3.3, welche im Bereich der optimalen Skalierung begründet sind und auf skalentechnische Probleme hinweisen. In erstgenanntem Kapitel wird auch eine erste Möglichkeit zur longitudinalen Analyse der Zufriedenheit vorgestellt. Kapiteln 3.3.4, 3.3.5 und 3.3.6 bauen aus drei unterschiedlichen Gedankenströmen heraus longitudinale Modelle auf. Kapitel 3.3.4 löst den Aufbau über die Einführung unterschiedlicher Levels, welche zuerst aus statischer Sicht besprochen werden, um unter anderem auch die im theoretischen Teil besprochenen nicht-linearen Effekte anhand neuerer Methoden zu durchleuchten. Schließlich wird der hierarchische Ansatz für longitudinale lineare und nicht-lineare Modelle übernommen. Kapitel 3.3.5 berücksichtigt die im Theorieteil angesprochene Messfehlerverbesserung über latente Konstrukte und gruppiert die Zufriedenheitsteilaspekte für die darauf folgenden beiden Kapitel. Kapitel 3.3.6 stellt eine Modellgruppe vor, welche variationsreiche Möglichkeiten zur Messung von Zufriedenheitsveränderungen zulässt. Am Ende des Kapitels wird schließlich auch eine simultane Schätzung der ebenfalls im Theorieteil angesprochenen Segmentierungsnotwendigkeit berücksichtigt. Die Modellgruppe des Kapitels 3.3.7 legt den longitudinalen Analyseschwerpunkt in Zufriedenheitstypologien und die Wahlwahrscheinlichkeiten und Übergänge der zugrundeliegenden Profile werden für die Respondenten ermittelt. Kapitel 4 listet abschließend Schlussfolgerungen und Verwertbarkeitsmöglichkeiten der gewonnenen Einblicke auf.

3. Ausgewählte Konzepte aus der Zufriedenheitsliteratur

Im Folgenden werden unterschiedliche Konstrukte der Zufriedenheits- und Wichtigkeitsmessung theoretisch dargestellt und anhand des Kinodatenatzes teilweise empirisch überprüft. Der dem Literaturteil zugrundeliegende meist chronologische Ablauf der Subthemen verdeutlicht die Entwicklung der unterschiedlichen Gedankenströme. Ergebnisse und damit einhergehende Kritik werden der theoretischen Vorstellung bzw. der empirischen Überprüfung angeschlossen.

3.1. Zufriedenheit und Wichtigkeit

3.1.1. Human Motivation

Einen einleitenden Rückblick zu einzelnen Punkten der Zufriedenheitsproblematik bildet hier die Motivationstheorie von Maslow (1943), welche eine fundamentale Argumentationsgrundlage zur Entstehung und Befriedigung einzelner Bedürfnisse darstellt. Da die Entstehung von Zufriedenheit aus der Bedürfnisbefriedigung erwächst, werden eben genau diese Kernaspekte methodisch diskutiert. Die Bedürfnistreiber benennt Maslow in der Reihenfolge ihres hierarchischen Auftretens, als *basic needs (physiological need), safety need, love need, esteem need* und *need for self-actualization*. Da die Bedürfnisse in einer Hierarchie aufgebaut sind, taucht ein Bedürfnis erst auf, wenn ein vorgelagertes zufriedengestellt wurde und ist somit immer von der Bedürfnisbefriedigung anderer Bedürfnisse abhängig. ‚Also no need or drive can be treated as if it were isolated or discrete; every drive is related to the state of satisfaction or dissatisfaction of other drives (Maslow, 1943).‘ In höheren Hierarchiestufen sinkt der prozentuelle Grad an Erfüllung und somit trifft man am Weg zur Spitze der Bedürfnispyramide auf steigende Unzufriedenheit.

Maslow (1943) verwendet bereits zum damaligen Zeitpunkt Argumente, welche die Streuung des Bedürfnislevels über die Population begründen. Das damalige Bedürfnislevel der Amerikaner wird wie folgt beschrieben: ‚The average American citizen is experiencing appetite rather than hunger when he says „I am hungry“‘. Er unterstellt die Suche nach Bedürfnisbefriedigung auf allen Ebenen der Hierarchie des Bedürfnisbaumes. Dass es auch Populationen gibt, welche auf unterschiedlichen Bedürfnisstufen stehen, steht wohl damals als auch heute außer Frage. In der Kinostudie befinden sich die befragten Personen auf Grund der Besuchsmöglichkeit eines Kinos auf einem höheren Bedürfnislevel und halten Ausschau nach der Befriedigung der höchsten Bedürfnisse. Somit verstehen sich die durchgeführten Untersuchungen als Messung für den Grad an

Erfüllung der Bedürfnisse und sämtliche Motivatoren spielen dabei eine Rolle. Zufriedenheit, als Ausdruck der bewerteten Performance, basiert in der vorliegenden Studie auf dem Level der erreichten Bedürfnisbefriedigung mit vorangehenden Erwartungen. Jedoch wird kein allumfassender Populationslevel unterstellt. Stattdessen spielt die Populationsmischung eine wesentliche Rolle, was in späteren Segmentierungslösungen berücksichtigt wird.

Auch die Wichtigkeit ist in der Zufriedenheitswelt von entscheidender Bedeutung. Maslow (1943) begründet den Zusammenhang zur Wichtigkeit bei der Suche nach Zufriedenheit folgendermaßen: ‚An act is psychologically important if it contributes directly to satisfaction of basic needs. The less directly it so contributes, or the weaker this contribution is, the less important this act must be conceived to be from the point of view of dynamic psychology.‘ Dieses Argument wird später genauer durchleuchtet, wenn untersucht wird, welche Serviceleistungen besondere Wichtigkeit erfahren und welche als weniger wichtig angesehen werden.

Wurde ein Bedürfnis über eine längere Zeitspanne zufriedenstellend befriedigt, so wird dieses neu bewertet. Daraus folgt eine dynamisch sinkende Wichtigkeit. Er schreibt, ‚... needs cease to play an active determining or organizing role as soon as they are gratified.‘, und erläutert diese Thematik beispielhaft. Menschen, die noch nie an Hunger gelitten haben, sehen dieses Bedürfnis als nicht wichtig an und halten Ausschau nach höheren Bedürfnissen, welche im Moment als wichtig betrachtet werden. Umgekehrt werden unwichtige Bedürfnisse nach einer zeitlichen Spanne der Entbehrung neu evaluiert. Beispielsweise gibt ein Arbeiter seinen Job auf um sein Selbstwertgefühl zu behalten. Aufgrund von Geldmangel steigt das Grundbedürfnis Hunger und er wird versuchen den Job zurückzubekommen, auch wenn dafür der Verlust seines Selbstwertgefühles in Kauf genommen werden muss. Diese Betrachtungen werden vielseitig argumentiert. Personen tolerieren eher einen Entzug, wenn das Bedürfnis immer befriedigt wurde. Musste jedoch in der Vergangenheit darauf verzichtet werden, fallen die Zufriedenheitsauswirkungen unterschiedlich aus. Wichtigkeiten dürfen nicht als statisches Konzept angesehen werden, sondern gelten so wie die Zufriedenheit als fortwährend dynamisch.

3.1.2. Importance – Performance

Die *Importance-Performance-Analyse* (IPA) von Martilla und James (1977) stellt den statischen Zusammenhang zwischen Zufriedenheit und Wichtigkeit dar. Die Kundenzufriedenheit wird als Funktion der Erwartungen bezogen auf die Wichtigkeit und die Beurteilung der Performance der Attribute beschrieben.

Zuerst wird nach der Wichtigkeit und der Zufriedenheit der Attribute gefragt. Im Anschluss werden die Mittelwerte beider Bewertungen für jedes Attribut errechnet und als Punkte in einem zweidimensionalen Grid eingezeichnet. Aufgrund der fehlenden Intervallskalierung sollte die Visualisierung über die Mediane führen. Falls diese zu ähnlichen Ergebnissen führt, wird die Mittelwertlösung aufgrund der zusätzlich enthaltenen Information empfohlen. Über den Gesamtmittelwert aller Attribute wird der zweidimensionale Raum in vier Quadranten unterteilt, welche zu unterschiedlichen Managementempfehlungen, wie in Diagramm 1 dargestellt, führen.

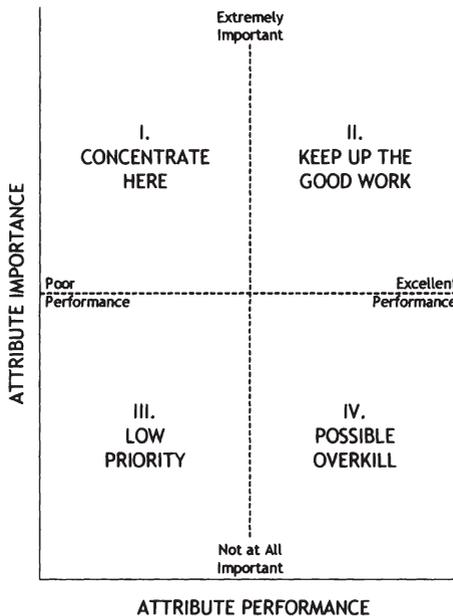


Diagramm 1: Action Grid (Crompton und Duray, 1985)

- I. *Concentrate here*: Das Attribut wird als wichtig betrachtet, die Performance als schlecht. Eine Erhöhung des Servicelevels bewirkt in diesem Quadranten maximale Verbesserung.
- II. *Keep up the good work*: Das Attribut ist wichtig und wird zur Zufriedenheit der Kunden erfüllt. Der Servicelevel sollte beibehalten werden.
- III. *Low priority*: Das Attribut ist weder wichtig, noch ist die Performance gut. Schlechte Performance stellt keinen Nachteil dar und kann, solange das Attribut nicht an Wichtigkeit gewinnt, auf gleichbleibendem Level beibehalten werden.

IV. *Possible overkill*: Das Attribut ist nicht wichtig, die Performance jedoch sehr gut. Möglicherweise wird das Anspruchslevel der Kunden übertroffen und somit nicht mehr in selbigem Ausmaß wahrgenommen. Die überschüssigen Ressourcen sollten auf andere Attribute umverteilt werden.

Es gibt unterschiedliche Möglichkeiten die Zufriedenheit der Kunden, welche ein Produkt kaufen bzw. eine Dienstleistung in Anspruch nehmen, zu verbessern. Entweder die Performance wird verbessert, oder die empfundene Importance gesenkt. Beide Alternativen führen zur Steigerung der empfundenen Servicequalität. Crompton und Duray (1985) schlagen Verbesserungen des Modells vor und beschreiben vier Importance-Performance Plots, deren Unterschiede zur Identifizierung eines bestbewährten Plots verwendet werden. Die direkte Abfrage der Wichtigkeit kann zu Verzerrungen führen. Deshalb wird anstatt der deskriptiven Lösung mit statistischen Hilfsmitteln gearbeitet, um die Wichtigkeit auf anderem Wege messbar zu machen.

In der ersten Methode werden die Pearson-Korrelationskoeffizienten der Wichtigkeits- und Performancewerte jedes Attributes berechnet. Anschließend werden die absoluten r-Werte und die Mediane der Wichtigkeitswerte absteigend gerangreicht. Danach werden diese auf Basis beider verbunden und gerangreicht und zusammen mit der Medianperformance geplottet. In der zweiten Methode werden die Korrelationswerte der Spearman's Rangordnungskorrelation mit den Medianwerten der Performancebewertung verwendet. Aufgrund der fehlenden Intervallskalenabstände wird ein nicht-parametrischer Korrelationskoeffizient verwendet. Die weiteren Variationen betreffen die Mittelwert- bzw. Medianauswahl der Wichtigkeits- bzw. Performancewerte. Ein Vergleich zwischen den unterschiedlichen Lösungen zeigt, dass sich die beiden Korrelationslösungen am meisten ähneln. Somit kann der Unterschied der parametrischen und der nicht-parametrischen Lösung nicht bestätigt werden. Die direkten Mittelwertberechnungslösungen sind ebenfalls sehr ähnlich. Die größten Unterschiede finden sich beim Vergleich zwischen direkter Befragung und statistischer Lösung. Der Grund wird dem Problem des Zusammenhanges zwischen Produktmerkmalen und Wahrnehmungen zugeschrieben. Am häufigsten treten unterschiedliche Zuordnungen zwischen Quadrant I und II, als auch II und IV, auf. Daraus können Fehleinschätzungen resultieren. Attribute weisen geringe Wichtigkeit auf, obwohl sie eine hohe Wichtigkeit und niedrige Performance haben. Ein Attribut wird als wichtig empfunden, wo die Performance schlecht ist, aber eigentlich könnten Einsparungen erzielt werden. Sie werten daher die Unterschiede zwischen der selbstbewerteten Wichtigkeit

und der indirekten Lösung als eine Verbesserung durch die Verwendung der statistischen Verfahren.

Auch etliche alternative Lösungen wurden bisher entwickelt. Beispielsweise werden die Rangwerte nach Berechnung und Reihung der Mittelwerte bzw. Medianwerte aufgetragen. Auch reine Performancemessungen mit anschließender Regression der Attributperformancebewertungen auf die Gesamtzufriedenheit werden durchgeführt. Als Wichtigkeitsmaß dienen dabei die standardisierten Regressionskoeffizienten. Auch im Bereich der Conjoint-Analyse finden sich Lösungen. Mit Hilfe einer Dummy-Variablen-Regression werden Koeffizienten für einzelne Attributlevels berechnet und die Spannen der Koeffizienten eines jeden Attributes dienen zur Bewertung der Wichtigkeit des Attributes. Möglichkeiten die Wichtigkeit zu identifizieren gibt es viele. Entscheidend ist jedoch der Zusammenhang zwischen Wichtigkeit und Zufriedenheit um Veränderungsauswirkungen der einen Eigenschaft auf die andere übertragen zu können.

Sampson und Showalter (1999) gehen deshalb genauer auf die Wirkungen zwischen Wichtigkeit und Zufriedenheit ein. Die Schwachpunkte der IPA sehen sie in der Unabhängigkeit der Wichtigkeit in Bezug auf die Performance. Wichtigkeit ist keine Punktschätzung, sondern eine Funktion, welche in kausalem Zusammenhang mit der Performance steht. Somit führen Veränderungen der Performance zu Veränderungen in der Wichtigkeit. In Verbindung mit der Performancebewertung steht die Priorisierung der Variablen für die Dringlichkeit der Verbesserung. Diagramm 2 zeigt den kontinuierlichen Übergang zwischen *overkill* und *urgent*. Dieser ist nicht durch horizontale und vertikale Trennungen beschreibbar, sondern durch einen trade-off zwischen Wichtigkeit und Zufriedenheit.

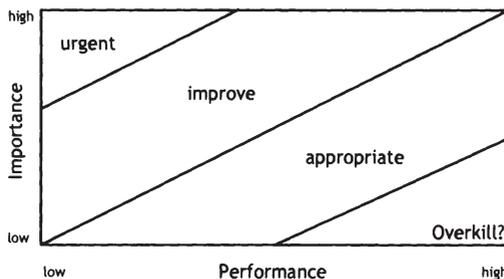


Diagramm 2: Slack's IPA Matrix (Sampson und Showalter, 1999)

Existiert eine signifikante Korrelation zwischen Wichtigkeit und Performance, so bedeutet dies, dass die beiden Konstrukte nicht orthogonal zueinander stehen. Laut Maslow's Bedürfnispyramide wird ein negativer Zusammenhang angenommen. Auch Herzberg argumentiert unter dem Begriff Hygienefaktoren auf diese Art und Weise. Wenn ein Bedürfnis nicht mehr länger befriedigt wird, also die Performance sinkt, so tritt die Wichtigkeit nach Befriedigung dieses Bedürfnisses zum Vorschein. Es wird eine verbesserte Fragestellung nach der Wichtigkeit vorgeschlagen, nämlich: „Wie wichtig ist Ihnen dieses Attribut unter gegebenem Zufriedenheitslevel?“.

Attribute, welche bei niedriger Performance sehr wichtig und bei hoher Performance unwichtig sind werden als *qualifier* oder *order-qualifying criteria* bezeichnet. *Order-winning criteria* sind jene Attribute, welche im Zuge der Kaufentscheidung die ausschlaggebenden Kriterien beim Vergleich der in Frage kommenden Güter oder Dienstleistungen bilden. Sie gewinnen an Wichtigkeit bei höherem Performancelevel. Ein Attribut mit hoher Wichtigkeit, bei hoher als auch bei niedriger Performance, ist sowohl ein *qualifier* als auch ein *order-winning criteria*. Die Ermittlung der Wichtigkeitswerte über eine Regression widerspricht jedoch den getroffenen Annahmen. Deshalb bevorzugen Sampson und Showalter (1999) eine direkte Befragung. Ein Vergleich unterschiedlicher Messmöglichkeiten (*direct rating*, *constant-sum scale* und *anchored scale*) ergab keine signifikanten Unterschiede und führt zur Wahl der *partial ranking*-Methode. Dabei werden die wichtigsten drei Attribute gereiht und die prozentuelle Größe des Auftretens als Wichtigkeitswert verwendet. Es werden drei Attribute vorgeschlagen, da weniger zu wenig Information bringen und mehr keine sinnvolle Wichtigkeitsdifferenzierung zulassen.

In Tabelle 1 werden die grundlegenden neuen Annahmen zusammengefasst. Es kann sowohl der postulierte Zusammenhang zwischen Wichtigkeit und Performance, als auch die negative Ausrichtung gezeigt werden. Die Kausaleigenschaft wird mit Hilfe der *performance-importance function* (PIF) überprüft. Eine Regression der Veränderungen der Performancebewertungen auf die Wichtigkeit zeigt einen signifikant negativen Einfluss. Verbesserungen der Performance führen zu sinkenden Wichtigkeitsbewertungen. Umgekehrt können Verbesserungen der Performance auch durch Veränderungen der Wichtigkeit erreicht werden.

Es wird weiters angeführt, dass die Konsumenten mit steigender Wichtigkeit empfindlicher gegenüber Veränderungen werden. Die Empfindlichkeit steigt somit auch bei sinkender Performance. Somit sollte eine Funktion, welche die

Attributs- mit der Gesamtpformance oder der Zufriedenheit verbindet auch steiler bei niedriger Performance, als bei höherer Performance sein.

Old Importance-Performance Assumption	New Assumption under these Hypotheses
Importance is sufficiently characterized as a point estimate, independent of performance.	Importance is best characterized as a function of performance.
Implementing improvement initiatives only moves attributes along the performance dimensions.	Improvement initiatives will likely have a simultaneous effect on performance and importance.
Every attribute that is highly important will be an 'order-winning criteria' at high performance.	Some attributes are highly important only under poor performance, and will never be 'order-winning criteria'.
Importance values can be derived by regressing attribute performance scores on an overall performance measure.	Since importance correlates with performance, statistical techniques that require independence are inappropriate (such as OLS regression).

Tabelle 1: Implications of hypotheses (Sampson und Showalter, 1999)

Auch Taguchi's *quality loss function* zeigt einen steileren Anstieg für jene Performance, welche weiter vom Idealpunkt entfernt ist. Weiters ist diese steiler für Verluste als für Gewinne. In vorangegangenen Studien konnte bereits gezeigt werden, dass die Zufriedenheit signifikant empfindlicher bei negativer Diskonfirmation (Performance liegt niedriger als die Erwartungen) als bei positiver Diskonfirmation (Erwartungen werden übertroffen) ist. Diese Phänomene visualisieren sie auf Basis empirischer Daten in Diagrammen 3.

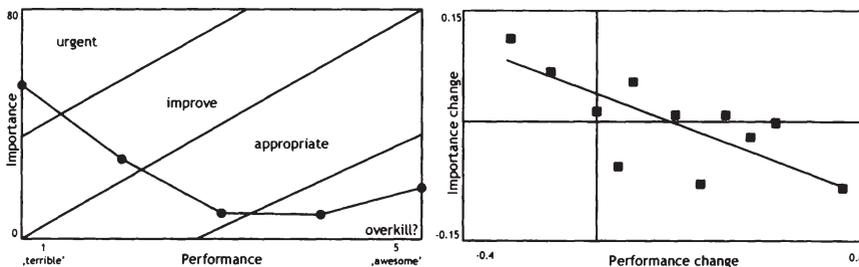


Diagramm 3: Importance Performance Zusammenhang (Sampson und Showalter, 1999)

Die Konvexität der Kurven, welche durch einen Anstieg der Wichtigkeit bei hoher Performance zustande kommt, wird durch die *zone of tolerance* erklärt. Die *zone of tolerance* ist ein Performance Bereich, in welchem der Kunde

zufrieden ist und innerhalb dessen kleinere Variationen möglich sind, ohne jedoch Einfluss auf die Wahrnehmung zu haben. Lediglich außerhalb dieses Bereiches wirken sich Veränderungen auf die empfundene Servicequalität aus. Der Konsument ist also unempfindlich gegenüber kleinen Veränderungen, da die Performance den Erwartungen auf gegebenem Level entspricht. Jene Spanne bzw. jener Punkt, welche zu minimaler Wichtigkeit führen, entsprechen deshalb der erwarteten Performance. Davon ausgehend beeinflusst Diskonfirmation den Konsumenten in beide Richtungen durch steigende Attributswichtigkeit. Die Breite der *zone of tolerance* kann deshalb aufgrund der Stärke der Konvexität der PIF-Funktion um den Erwartungspunkt festgestellt werden. Hat ein Item keinen Minimalpunkt, so wird angenommen, dass die Erwartungen an der positiven Spitze der Performance Skala liegen und es ist unmöglich, die Erwartungen zu übertreffen.

Konsumenten beurteilen die Produkte anhand eines begrenzten Sets an Attributen, wobei manche bezüglich der Beeinflussung der Zufriedenheit wichtig sind und andere nicht. Die Grenzen der *zone of tolerance* sind jene beiden Punkte, ab der Diskonfirmation vorliegt und je nachdem ob positive oder negative, dies auch unterschiedliche Auswirkungen im Zufriedenheitsbereich hat.

3.1.3. Eindimensional vs. zweidimensional

Die wohl bekannteste Methode formulieren Kano et al. (1984). Die Idee dazu stammt aus dem Zusammenhang zwischen objektiver Qualität, physischer Erfüllung und resultierender subjektiver Qualität, ausgedrückt durch die Kundenzufriedenheit. Die Korrelation zwischen beiden sagt etwas über die Gesamtproduktqualität aus und nicht nur über die einzelnen Qualitätselemente. Sie entfernen sich von der eindimensionalen Sichtweise, hin zu einer zweidimensionalen, wie in Diagramm 4 dargestellt.

Kano et al. (1984) beschreiben die auffindbaren Qualitätselemente wie folgt, wobei die drei erstgenannten die relevantesten sind:

'Attractive quality elements: Quality elements that when fulfilled provide satisfaction but when not fulfilled are acceptable. ...

One-dimensional quality elements: Quality elements that result in satisfaction when fulfilled and in dissatisfaction when not fulfilled. ... the conventional way of recognizing quality. ...

Must-be quality elements: Quality elements that are absolutely expected (taken for granted when fulfilled) but result in dissatisfaction when not fulfilled. ...

Indifferent quality elements: Quality elements that neither result in satisfaction nor dissatisfaction, regardless of whether they are fulfilled or not. ...

Reverse quality elements: Quality elements that result in dissatisfaction when fulfilled and satisfaction when not fulfilled.'

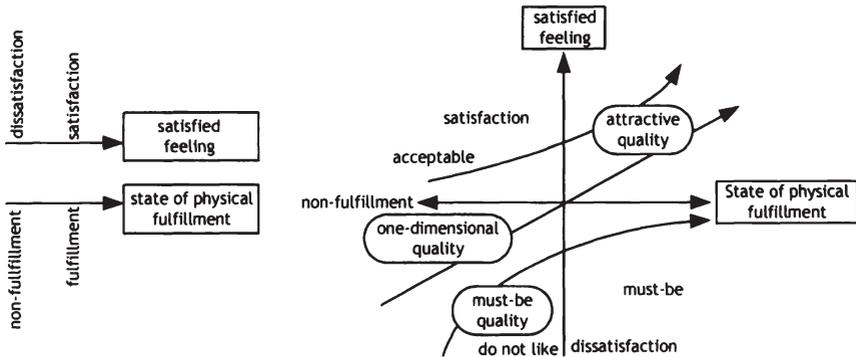


Diagramm 4: One-dimensional vs. two-dimensional recognition method (Kano et al., 1984)

Durch die Klassifizierung der Servicequalitätselemente in die genannten Kategorien mit Hilfe der Kano-Methode wird die Validität einer zwei-dimensionalen Betrachtungsweise verfolgt. Für jedes Attribut wird die Häufigkeit der Zuordnung zu einer der Kategorien bestimmt. Weisen alle Attribute die Hauptaussprägung *one-dimensional quality elements* auf, so kann Eindimensionalität (Diagramm 4, linke Graphik) angenommen werden. Dies ist in der Originalstudie nicht der Fall. Deshalb wird Eindimensionalität verworfen und Zweidimensionalität angenommen (Diagramm 4, rechte Graphik).

Werden Qualitätselemente im Bereich *attractive* gemessen, so ist die Tendenz Richtung *indifferent*, oder auch *reverse*, stärker verglichen mit anderen Elementen. Dies bedeutet, dass Qualitätselemente für viele wünschenswert sind, gleichzeitig aber nicht für andere Personen. Diese Qualitätselemente widerspiegeln ebenfalls die Nutzerdiversität. Kano et al. (1984) appellieren, diese Unterschiede nach demographischen, etc. Eigenschaften zu untersuchen.

3.1.4. Satisfier - Dissatisfier

Es wird folgend davon ausgegangen, dass es Attribute gibt, welche lediglich Einfluss auf die Zufriedenheit haben und umgekehrt jene, welche lediglich Einfluss auf die Unzufriedenheit haben. Eine Methode derartige Servicequalitätsdimensionen bzw. -eigenschaften aufzudecken stellt Johnston (1995)

vor. Im Zuge der *critical incident technique* (CIT) bewerten Konsumenten ihr Empfinden mit den Serviceleistungsattributen als *satisfying factors* bzw. *dissatisfying factors*, jedoch mit dem Nachteil, dass nur extreme Bewertungen abgegeben werden und Qualitätskriterien in der Nähe oder innerhalb der Toleranzzone weitgehend ausgeschlossen werden.

Eine nicht signifikante Spearman's Rangkorrelation der relativen Häufigkeiten, wie oft die jeweiligen Servicekriterien als *satisfier* bzw. *dissatisfier* angegeben werden, zeigt, dass ein Unterschied zwischen den beiden Zufriedenheitseigenschaften mit einer Wahrscheinlichkeit über 0,90 besteht. Der Korrelationskoeffizient liegt bei etwa 0,1 und so kann ein gegenteiliger Zusammenhang ausgeschlossen werden. Für vier der ermittelten Kriterien kann eine eindeutige Klassifizierung zu einer der beiden Kategorien festgestellt werden. ‚Integrity‘ und ‚aesthetics‘ fallen in die Kategorie *Dissatisfaction*, ‚commitment‘ und ‚cleanliness‘ in die Kategorie *Satisfaction*, wie in Diagramm 5 zu sehen ist. Die restlichen Attribute üben auf beiden Seiten Einfluss aus. Es ist jedoch erkennbar, dass einige der Attribute vermehrt auf der *Satisfaction*-Seite und weniger auf der *Dissatisfaction*-Seite auftauchen bzw. umgekehrt. Diese Attribute haben somit auf einer der beiden Seiten einen größeren Einfluss. Je gleichmäßiger die Häufigkeiten auf beiden Seiten verteilt sind, desto eher spielt dieses Attribut für beide Zufriedenheitsarten eine Rolle. Die Ergebnisse ermöglichen einen Einblick darin, für welche Faktoren Zufriedenheit gesteigert bzw. Unzufriedenheit gesenkt werden muss. Empfohlen wird die Anzahl an zufriedenstellenden Erfahrungen zu erhöhen. Der Mangel an derartigen Attributen trägt zwar nicht zur Unzufriedenheit bei, das Vorhandensein kann jedoch zur Stufe *delight* führen.

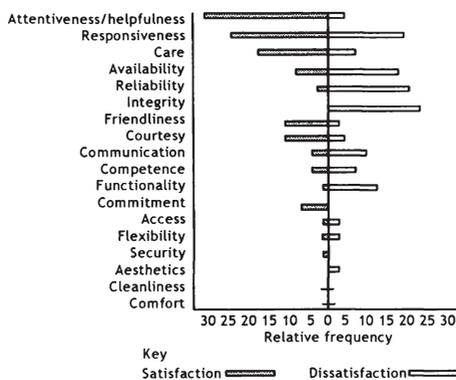


Diagramm 5: Satisfaction/dissatisfaction Kriterien anhand der Critical Incident Technique (Johnston, 1995)

Johnston (1995) spricht auch den Bezug zu den Qualitätsdimensionen an. Die Klassifizierungsergebnisse können dazu verwendet werden, die Attribute nach deren Auswirkung bezogen auf das SERVQUAL-Modell hin, zu untersuchen. Darauf wird hier nicht speziell eingegangen, um den Fokus der Arbeit auf die Zufriedenheitstypeneinteilung beizubehalten.

3.1.5. Kano-Modell

Matzler et al. (1996) analysieren nach Vorbild des Kano-Modells genauer die Eigenschaften bzw. Hintergründe der *delight*-Faktoren. Sie stellen sich die Frage, welche Produkte oder Dienstleistungen zu hoher Konsumentenzufriedenheit führen, welche Produktfeatures die Zufriedenheit stärker als proportional beeinflussen und welche Attribute ein absolutes Muss für die Konsumenten sind. Bis dato wurde davon ausgegangen, dass die Steigerung der empfundenen Qualität immer mit einer Steigerung der Zufriedenheit einhergeht, und umgekehrt. Dies kann jedoch nicht für *must-be* Faktoren angenommen werden. Diese müssen erfüllt werden, damit Zufriedenheit überhaupt erst entstehen kann bzw. bewirken bei Übererfüllung der Qualitätsempfindung keinen weiteren Zufriedenheitsanstieg. Die drei Attributeigenschaften *must-be*, *one-dimensional* und *attractive requirements* werden in Diagramm 6 verdeutlicht. *Must-be requirements* müssen erfüllt werden um Zufriedenheit überhaupt zuzulassen. Sie stellen eine vorangehende Bedingung dar. Ihre Erfüllung hat lediglich die Vermeidung der Unzufriedenheit zum Ziel. *One-dimensional requirements* beschreiben Attribute, deren Verbesserung der Eigenschaften bzw. Erhöhung der Erfüllung automatisch zu einem Anstieg der Zufriedenheit führt. Umgekehrt führt ein Sinken der Qualität zu einer abgeschwächten Zufriedenheit. *Attractive requirements* üben den stärksten Einfluss auf die Zufriedenheit aus. Sie führen zu überproportionalem Anstieg der Zufriedenheit, wirken sich aber bei Nichterfüllung nicht auf die Unzufriedenheit aus. Die Graphik wird um die konkreteren Eigenschaften der Faktoren erweitert.

Aus dieser Klassifizierung können verschiedene Handlungsempfehlungen abgeleitet werden. ‚Must-be requirements‘ stellen eine Marktbarriere dar. Ohne deren Erfüllung wird ein Produkt- oder Serviceangebot vom Kunden nicht in Betracht gezogen. Werden diese jedoch bereits zu Genüge erfüllt, so ist eine weitere Investition in diese Attribute ineffizient, da sie zu keinem weiteren Anstieg der Zufriedenheit führen wird. *One-dimensional* und *attractive requirements* hingegen sind jene Eigenschaften, welche gefördert und verbessert werden müssen. Deren Einflussunterschiede werden ausschlaggebend dafür sein, auf welche Eigenschaften das Hauptaugenmerk gelegt werden muss. Soll beispielsweise ein Produkt von der Konkurrenz abgehoben werden, so wird dies

am einfachsten über die *attractive requirements* erreichbar sein. Haben bereits alle Marktteilnehmer diese Eigenschaft in ihr Produkt/Service inkludiert, so hat die Eigenschaft keinen überproportionalen Einfluss mehr auf die Zufriedenheit. Das Vorhandensein wird als Standard verstanden.

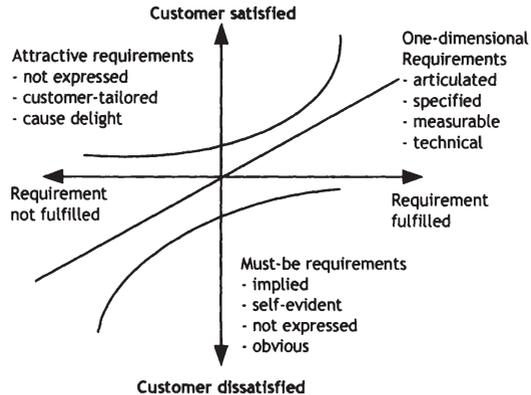


Diagramm 6: Kano's model of customer satisfaction (Matzler et al., 1996)

Auch den zeitlichen Aspekt erwähnen Matzler et al. (1996). Betrachtet man die unterschiedlichen Attributseigenschaften nämlich über die Zeit hinweg, so können *attractive requirements* irgendwann zu *one-dimensional requirements* und dann zu *must-be requirements* werden. Wird ein Service immer wieder auf gleichbleibendem Level erfüllt, so verliert es im Zeitablauf an ihrem Zufriedenheit stiftenden Nutzen und wird als gegeben vorausgesetzt.

Im Folgenden wird die Vorgehensweise der Kano-Methode im Detail beschrieben. Zuerst müssen die Eigenschaften identifiziert werden. Mit Hilfe qualitativer Interviews können zwar die aus Konsumentensicht relevanten Eigenschaften gefunden werden, *attractive requirements*, welche oftmals unerwartet als Zusatzservice angeboten werden, werden jedoch in derartigen Fokusgruppeninterviews nicht zur Sprache kommen. Deshalb wird die Befragungsweise adaptiert, 'Which associations does the customer make when using product x?, Which problems/defects/complaints does the customer associate with the use of product x?, Which criteria does the customer take into consideration when buying product x?, Which new features or services would better meet the expectations of the customer? What would the customer change in product x?'. Mit Hilfe dieser Fragen können die unterschiedlichen *requirements* besser ermittelt werden. Im Anschluss werden die Probanden nach den aufgefundenen Eigenschaften auf zwei unterschiedliche Arten befragt. In

Tabelle 2 misst die erste positiv formulierte Frage (functional) die Reaktion des Konsumenten, wenn das Produkt die Eigenschaft hat, in der zweiten negativ formulierten, wenn es die Eigenschaft nicht hat (dysfunctional) (Matzler & Hinterhuber, 1998).

Frage	Antwort
<i>Functional form</i>	(1) I like it that way
If the edges of your skis grip well on hard snow, how do you feel?	(2) It must be that way
<i>Dysfunctional form</i>	(3) I am neutral
If the edges of your skis do not grip well on hard snow, how do you feel?	(4) I can live with it that way
	(5) I dislike it that way

Tabelle 2: Functional/dysfunctional Form (Matzler et al., 1996)

Aufgrund der Beantwortung werden die Eigenschaften in Tabelle 3 in sechs unterschiedliche Attributsklassen unterteilt.

Functional (positive) question	Dysfunctional (negative) question				
	(1) Like	(2) Must-be	(3) Neutral	(4) Live with	(5) Dislike
(1) Like	Q	A	A	A	O
(2) Must-be	R	I	I	I	M
(3) Neutral	R	I	I	I	M
(4) Live with	R	I	I	I	M
(5) Dislike	R	R	R	R	Q

Note: A: attractive; M: must-be; R: reverse; O: one-dimensional; Q: questionable; I: indifferent

Tabelle 3: Einteilung der Eigenschaftsausprägung (Matzler et al., 1996)

Die zusätzlich auftauchenden Eigenschaften können wie folgt erklärt werden: *questionable* - Eine richtige Beantwortung ist fraglich; *indifferent* - Die Eigenschaft ist dem Befragten gleichgültig und er wird dafür nicht mehr Geld ausgeben; *reverse* - Es wird das Gegenteil der momentanen Produkt- bzw. Serviceeigenschaft erwartet. Jene Faktorbeschreibung, welche prozentuell am häufigsten zutrifft, bestimmt die Eigenschaft des Attributes. Eine erste Interpretation kann aus den Häufigkeiten der sechs Eigenschaften für die einzelnen Attribute folgen. Diese können zwischen Segmenten variieren, da unterschiedliche Konsumenten andere Bedürfnisse aufweisen können. Personen, die einen gewissen Status gewohnt sind, können ein Attribut als Standard betrachten. Andere wiederum, welche mit diesem Attribut weniger Erfahrung aufweisen, können dieses als *one-dimensional* oder sogar als *attractive*

requirement klassifizieren. Eine Reihung der Eigenschaften erfolgt nach folgender Regel: $M > O > A > I$ Zuerst müssen die Mindestbedingungen erfüllt werden, um Unzufriedenheit zu vermeiden. Anschließend fällt die Konzentration auf die *one-dimensional* und *attractive requirements*. Mit einer einhergehenden Wichtigkeitsbefragung können die Attribute je nach Probandenempfinden ausgewählt werden.

Der *Customer Satisfaction-coefficient* (CS-coefficient) in Formel 1 bietet eine Interpretationshilfe. Die Werte schwanken zwischen 0 und 1 (*satisfaction*) bzw. -1 (*dissatisfaction*). Je näher der Wert bei 1 bzw. -1 desto stärker der Einfluss des Attributes auf die Zufriedenheit bzw. Unzufriedenheit, je näher bei 0, desto geringer. Die beiden Werte können als x- und y-Achse aufgetragen und interpretiert werden.

$$\text{Extent of satisfaction: } \frac{A + O}{A + O + M + I}$$

$$\text{Extent of dissatisfaction: } \frac{O + M}{(A + O + M + I) \times (-1)}$$

Formel 1: Customer Satisfaction-Coefficient (Matzler et al., 1996)

Um einen Vergleich zur Konkurrenz herzustellen wird vorgeschlagen, das eigene Produkt und das des Konkurrenten bewerten zu lassen. Daraus wird die Differenz gebildet und mit der befragten Wichtigkeitsbewertung multipliziert. Ein hoher negativer Wert zeigt einen Nachteil gegenüber der Konkurrenz, ein hoher positiver Wert einen Vorteil. Durch die Identifizierung der Attribute wird klar, dass Investitionen in *must-be*-Faktoren, welche sich auf einem zufriedenstellenden Level befinden, nicht sinnvoll sind. Es ist besser in *one-dimensional* oder *attractive requirements* zu investieren, da diese einen größeren Einfluss auf die empfundene Qualität haben und somit auch auf den Zufriedenheitslevel der Konsumenten.

Matzler und Hinterhuber (1998) erklären anhand einer empirischen Studie mit selbigem Schema mit den Kano-Faktoren die Loyalität. Sie postulieren einen nicht-linearen Zusammenhang zwischen Loyalität und Zufriedenheit. In extrem hohen Zufriedenheitsbereichen steigt die Loyalität überproportional, *zone of affection*, und in extrem niedrigen Zufriedenheitsbereichen sinkt die Loyalität überproportional, *zone of defection*. Dazwischen liegt die *zone of indifference*.

Lee et al. (2006) führen ebenfalls eine Kano-Studie durch, untersuchen die Unterschiede zwischen Wichtigkeit und Zufriedenheit und leiten Empfehlungen aus der *Importance-Performance* Matrix auf die zukünftigen Konkurrenzbeziehungen ab. Zuvor grenzen sie die Begriffe Kundenzufriedenheit und Servicequalität ab. Zufriedenheit wird als Resultat von Erwartung und Wahrnehmung gesehen. Auch vom psychologischen Status der Diskonfirmation der Erwartungen und den anfänglichen Erwartungen wird ausgegangen. So ist Zufriedenheit das psychologische Gefühl der aktuellen Erfahrung und des emotionalen Status im Anschluss an die Erfahrung. Servicequalität steht für den Fit zwischen Performance und den Qualitätsspezifikationen. Sie wird auch als Zufriedenheit mit dem Service angesehen, welche vom Unterschied zwischen erwartetem und wahrgenommenem Service abhängt. Weiters wird nicht nur die Performance, sondern auch der Serviceprozess miteinbezogen.

Auch die Unterschiede zwischen der eindimensionalen und der zweidimensionalen Sichtweise beschreiben Chen et al. (2006), als auch Lee et al. (2006). Das eindimensionale Qualitätsmodell geht von zufriedenen Konsumenten aus, also wenn die Erwartungen ausreichend erfüllt werden. Werden diese nicht erfüllt, sind sie unzufrieden. Im zweidimensionalen Qualitätsmodell, wie im Kano-Modell, führt die Hinlänglichkeit nicht zur Zufriedenheit, wie auch in der Motivator-Hygiene Theorie von Herzberg. Es existieren unterschiedliche Qualitätselemente, nämlich *attractive quality* und *must-be quality*, was zu einer Unterscheidung der Servicequalität in *attractive quality elements*, *one-dimension quality elements*, *must-be quality elements*, *indifferent quality elements* and *reverse quality elements* führt. Identifizierungsprobleme führen jedoch zu einer prozentuellen Zuordnung zu mehreren Faktoren. Aufgrund des Fehlens einer ausgewogenen Verteilung, empfehlen Lee et al. (2006) in das Kano Modell ein Fuzzy-Konzept einzubauen.

3.1.6. Asymmetrischer Einfluss

Die multiattributive Anwendung begründen Mittal et al. (1998) aufgrund der einfacheren Evaluierung der Performance auf Attributlevel verglichen mit der Gesamtzufriedenheit. Sie bietet die Möglichkeit Attribute aufzudecken, mit welchen der Konsument zufrieden ist, als auch jene, mit welchen dieser unzufrieden ist. Der Einfluss der Performance von Teilattributen auf die Gesamtzufriedenheit wird meist mit Hilfe der multiplen Regression analysiert. Diese unterstellt jedoch eine symmetrische und lineare Beziehung zwischen der Performance auf Attributlevel und dem abhängigen Konstrukt, wie beispielsweise Gesamtzufriedenheit oder Kaufabsicht. So leiten sie theoretisch und empirisch die Existenz einer asymmetrischen und nicht-linearen Beziehung

her. Die Asymmetrie begründen sie dadurch, dass eine Einheit an negativer Performance eines Attributes einen größeren Effekt auf das abhängige Konstrukt hat, verglichen mit einer gleichen Einheit an positiver Performance. So kann die negative Performance eines einzelnen Attributes die positive Performance mehrerer Attribute aufwiegen. Die Begründung durch die *Prospect Theory* wird in Diagramm 7 gezeigt. Die Referenzabhängigkeit sieht die Werteträger als Gewinne und Verluste in Bezug auf einen Referenzpunkt. Aufgrund der *loss aversion* ist die Funktion im negativen Bereich steiler, verglichen mit dem positiven Bereich. Durch die *diminishing sensitivity* sinken die Grenzwerte der Gewinne und Verluste mit deren Größe.

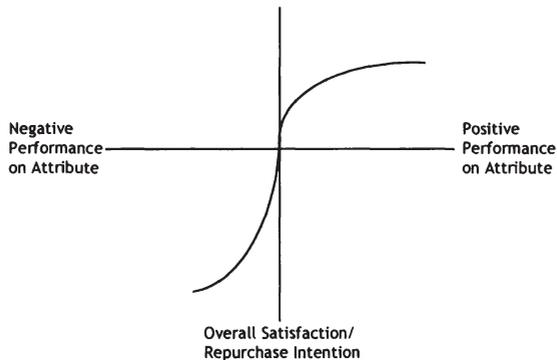


Diagramm 7: Asymmetric impact of attribute-level performance (Mittal et al., 1998)

Die Behauptung, dass negative Performance mit einem Attribut, einen größeren Einfluss auf die Gesamtsatisfaction hat, verglichen mit positiver Performance mit dem gleichen Attribut, kann durch höhere Regressionskoeffizienten der negativen Attribute, verglichen mit den niedrigeren Regressionskoeffizienten der positiven Attribute gezeigt werden. Auch für die Befragungsart der *Confirmation/Disconfirmation-Methode* (CD-Methode) kann diese Eigenschaft gezeigt werden.

Mit den logarithmierten Performannewerten wird gezeigt, dass auf höheren Levels positiver oder negativer Performance eines Attributes die Gesamtsatisfaction weniger beeinflusst wird, verglichen mit dazwischen liegenden Levels an Performance mit einem Attribut. Auch kumulativ gesehen, haben Attribute mit negativer Performance einen größeren Einfluss auf die Gesamtsatisfaction, verglichen mit Attributen positiver Performance. Jede Veränderung positiver oder negativer Performance hat einen kleineren Einfluss auf die Gesamtsatisfaction.

Der asymmetrisch kumulative Einfluss der positiven und negativen Performance auf die Gesamtzufriedenheit wird in Formel 2 mit $b1$ bzw. $b3$ berechnet. POSITIVE und NEGATIVE sind die Summen der positiven bzw. negativen Bewertungen der Attribute. Die *diminishing sensitivity* wird mit deren Quadraten mit in das Modell aufgenommen. $b1$ und $b2$ stehen für *diminishing sensitivity* für die positive Performance, $b3$ und $b4$ für die negative Performance. $b3$ ist größer als $b1$ und der Effekt kann gezeigt werden. Die Koeffizienten der quadrierten Attribute sind ebenfalls signifikant. Die Annahme kann teilweise gezeigt werden, da *diminishing sensitivity* nur für positive, nicht aber für negative Attribute gefunden wurde.

$$\text{Overall satisfaction} = b1 * (\text{POSITIVE}) + b2 * (\text{POSITIVE})^2 + b3 * (\text{NEGATIVE}) + b4 * (\text{NEGATIVE})^2$$

Formel 2: Diminishing sensitivity (Mittal et al., 1998)

Für viele Attribute hat die positive Diskonfirmation keinen signifikanten Einfluss. Somit gibt es keinen zusätzlichen Nutzen durch ein Übertreffen der Kundenerwartungen, wobei jedoch die negativen Konsequenzen, wenn die Erwartungen nicht erfüllt werden, groß sind. Weiters haben die Attribute unterschiedlichen Einfluss auf die Gesamtzufriedenheit. Sie werden Kern- bzw. Unterstützungsattribute auf Basis der Asymmetriestärke eingeteilt. Nutzen-erhaltende Attribute besitzen mehr Potential für negative Diskonfirmation. Nutzensteigernde Attribute haben ein größeres Potential für positive Diskonfirmation. Der Performance- und Diskonfirmationslevel der Vergangenheit spielt dabei auch eine Rolle. Wurde eine Erwartung immer erfüllt und nun plötzlich nicht mehr, so wird die Gesamtzufriedenheit beeinflusst. Lag die Erfüllung immer unter den Erwartungen, so hat eine neuerliche negative Performance geringen Einfluss, eine hohe Qualität jedoch einen bedeutenden.

3.1.7. Nicht-linearer Zusammenhang

Anderson und Mittal (1998) untersuchen asymmetrische nicht-lineare Zusammenhänge zwischen Performance und Zufriedenheit, Zufriedenheit und Wiederkaufsrate, und Wiederkaufsrate und Profit. Asymmetrie, teilen sie in positive und in negative Asymmetrie, Nichtlinearität kann sich ebenso positiv oder negativ auswirken. Die Behauptung wird aufgestellt, dass der Einfluss einer Steigerung unterschiedlich der einer gleich großen Senkung ist. Dies gilt nicht nur in Bezug auf die Richtung, sondern auch in Bezug auf die Größe. Ist der Einfluss einer Senkung größer, so spricht man von negativer Asymmetrie. Können Performancegewinne die Verluste aufwiegen, so spricht man von positiver Asymmetrie. Nichtlinearität tritt in der Form von *diminishing returns*

auf. So hat jede weitere Inputeinheit einen geringeren Einfluss als die vorangehende Einheit. Selbiger Einfluss existiert auch für steigende Erträge. Links in Diagramm 8 ist der traditionelle symmetrisch lineare Ansatz zu sehen. Die Veränderung einer Einheit Performance führt zur gleichen Veränderung der Gesamtzufriedenheit, gleich ob die Veränderung im Mittelbereich oder an den Extremen erfolgt.

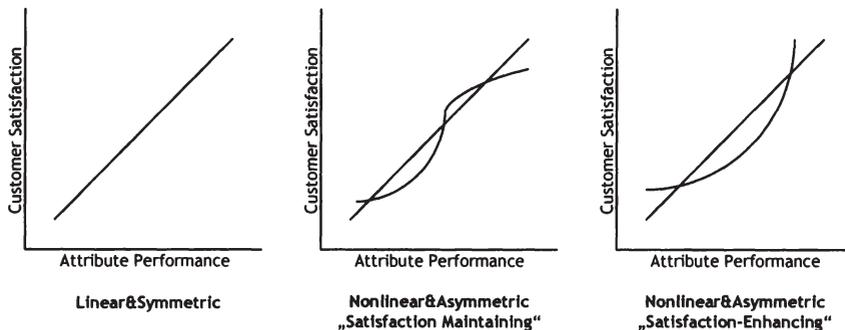


Diagramm 8: Performance-Satisfaction Link (Anderson und Mittal, 1998)

Das Bild in der Mitte beschreibt die *satisfaction-maintaining*-Attribute. Diese zeigen normalerweise negative Asymmetrie und sinkende Erträge beim Zufriedenheitseinfluss. Das Bild rechts zeigt positive Asymmetrie. Performanceveränderungen zum oberen Ende der Skala hin sind wichtiger als die im Mittelbereich, auch *surprise*, *delight* oder *satisfaction enhancing* genannt. Derartige Attribute sind meist unerwartet. Würde ein linearer symmetrischer Zusammenhang für *satisfaction-maintaining*-Attribute angenommen werden, so überschätzt der lineare Ansatz den Einfluss an den Extremen und unterschätzt diesen näher am Referenzpunkt der Konsumenten in der Mitte der Performance-skala. Die negative Asymmetrie führt zu größeren Fehlern im negativen Performancebereich. Der lineare Ansatz unterschätzt bei *satisfaction-enhancing*-Attributen den Einfluss an den Extremen. Durch die positive Asymmetrie ist der Fehler größer im positiven Bereich. Diese Argumente werden in Tabelle 4 zusammengefasst.

		Attribute Type	
		<i>Satisfaction-Maintaining</i>	<i>Satisfaction-Enhancing</i>
Attribute Performance	<i>Strong</i>	Overestimates Importance	Underestimates Importance
	<i>Weak</i>	Underestimates Importance	Overestimates Importance

Tabelle 4: Fehleinschätzung des linearen symmetrischen Satisfaction-Performance Links I (Anderson und Mittal, 1998)

Auch in der *Importance-Performance-Matrix* führt die Nichtbeachtung des asymmetrisch nicht-linearen Verlaufes, wie in Tabelle 5, zu Fehlern. Die Wichtigkeit eines Attributes basiert auf der Stärke der Beziehung zwischen Attributperformance und Zufriedenheit. Je stärker die Beziehung, desto wichtiger das Attribut, dargestellt durch die Steigung des Performance-Satisfaction Links. Je steiler der Anstieg, desto stärker die Beziehung und desto wichtiger das Attribut. Je flacher die Steigung, desto schwächer die Beziehung und desto weniger wichtig das Attribut.

		Attribute Importance				Attribute Importance	
		Low	High			Low	High
Attribute Performance	Strong	BACK OFF Enhancing attributes	KEEP UP Maintaining attributes	Attribute Performance	Strong	BACK OFF Maintaining attributes	KEEP UP Enhancing attributes
	Weak	IGNORE Maintaining attributes	IMPROVE Enhancing attributes		Weak	IGNORE Enhancing attributes	IMPROVE Maintaining attributes

Observed correction matrix based on using a linear&symmetric approach

Actual correction matrix based on using a nonlinear&asymmetric approach

Tabelle 5: Fehleinschätzung des linearen symmetrischen Satisfaction-Performance Links II (Anderson und Mittal, 1998)

Tabelle 5 weist auf folgende Fehler hin. Die linke Matrix empfiehlt die Verbesserung von *satisfaction-enhancing*-Attributen mit niedrigen Werten. Rechte Matrix schenkt diesen wenig Aufmerksamkeit wie aus der mittleren Graphik in Diagramm 8 hervorgeht. Niedrige Werte zu steigern führt zu keinen großen Zufriedenheitsveränderungen, da in diesem Bereich das Attribut nicht wichtig ist, wie die Flachheit des *Performance-Satisfaction*-Links zeigt. Weiters empfiehlt linke Matrix die Unterstützungssenkung der *satisfaction-enhancing*-Attribute mit hohen Werten, wobei rechte Matrix vom Gegenteil ausgeht. Die Erklärung findet sich in der rechten Graphik in Diagramm 8. Die Steigung ist größer im hohen Performancebereich und Performancesteigerungen führen zu großen Gewinnen in der Zufriedenheit. Wird die Nichtlinearität miteinbezogen, folgen weitere Fehler aus linker Matrix. Die Wichtigkeit hängt von der Performance ab. Obwohl *satisfaction-enhancing*-Attribute mit niedrigen Werten nicht wichtig sind, ist ihr Einfluss groß, sobald die Performance einen hohen Level erreicht. *Satisfaction-maintaining*-Attribute mit hohen Werten wirken unwichtig. Werden diese vernachlässigt, so führt dies zu einem Bereich, wo kleine Performanceänderungen zu großen Zufriedenheitsänderungen führen.

Zusammengefasst werden laut linker Matrix die Ressourcen den niedrig bewerteten *satisfaction-enhancing*-Attributen und den hoch bewerteten *satisfaction-maintaining*-Attributen zugesprochen, wobei hoch bewertete *satisfaction-enhancing*-Attribute an Unterstützung verlieren. Rechte Matrix empfiehlt das Gegenteil. Ressourcen sollten die Performance niedrig bewerteter *satisfaction-maintaining*-Attribute unterstützen und die Performance der hoch bewerteten *satisfaction-enhancing*-Attribute aufrechterhalten, wobei die Unterstützung der hoch bewerteten *satisfaction-maintaining*-Attribute gesenkt werden sollte.

Tabelle 6 bietet eine zusätzliche Handlungsempfehlungsmöglichkeit. In der Top-2-box steht der Prozentanteil der Personen, die mit 6/7 'extrem zufrieden' geantwortet haben, die Bottom-2-box betrifft die Antworten 1/2 'extrem unzufrieden'. *Satisfaction-enhancing*-Attribute mit niedrigen Top-2-box Werten und *satisfaction-preserving*-Attribute mit hohen Bottom-2-box Werten erhalten Priorität. Für erstere wird die positive Performance gesteigert, für letztere eine *zero-tolerance* gefunden. Normalerweise wird empfohlen die Performance wichtiger Attribute zu steigern. Die Asymmetrie appelliert zusätzlich zuerst die Abschwächung negativer Performance für *satisfaction-maintaining* Attribute, da diese folgenreicher ist, als die spätere Steigerung der Performance im positiven Bereich. Verglichen mit der Nicht-Linearität wird der Asymmetrie die größere Bedeutung beigemessen. Langfristig muss jedoch bei *satisfaction-maintaining* Attributen der sinkende Deckungsbeitrag bei höheren Performanceratings beachtet werden.

Anderson und Mittal (1998) erwähnen auch die Notwendigkeit die Wichtigkeitseinstellungen einzelner Attribute in Segmente zu gliedern und die zeitliche Veränderung der Zufriedenheitseigenschaften zu untersuchen: 'Whether an attribute is satisfaction maintaining or enhancing may change over time. ... in the initial introduction stage, an attribute is typically satisfaction enhancing ... Over time, however, as competitors match these attributes, they no longer surprise customers. Customers take them for granted, and they are likely to migrate into the satisfaction maintenance category. ... Note that industry structure is likely to affect the speed with which attributes migrate from the enhancement category to the maintenance category. ... Second, whether an attribute is satisfaction enhancing or satisfaction preserving may also depend on the length of the relationship between the customer and the product or service in question.'

Asymmetric key driver analysis: satisfaction with mutual fund adviser – Performance

Importance Analysis

Attribute	Top-2-box score (% rating 6/7)	Bottom-2-box score (% RATING 1/2)	Action priorities
Can solve problems in one call/visit	25%	35%	Satisfaction maintaining attribute with high dissatisfaction. Decrease bottom-2-box score
Clearly communicates required information	55%	5%	Satisfaction maintaining attribute with low dissatisfaction. Do nothing.
Have confidence in advice given by advisor	31%	10%	Satisfaction maintaining & enhancing attribute. If resources permit, increase top-2-box score.
Makes you feel like a valued customer	22%	2%	Satisfaction enhancing attribute with high satisfaction. If resource permit, increase top-2-box score.
Provides right amount of information & assistance	65%	16%	Satisfaction enhancing attribute with high satisfaction. Do nothing.

Tabelle 6: Action-steps chart (Anderson und Mittal, 1998)

3.1.8. Teilzufriedenheit - Gesamtzufriedenheit

Audrian-Pontevia (2006) beschreibt die Unterschiede zwischen der ein- und mehrdimensionalen Sichtweise. Anhand von Theorien nach Herzberg, Swan und Combs, oder Mausner und Snyderman's bifaktorieller Theorie, wird die Existenz von Attributen angenommen, welche zu steigender Gesamtzufriedenheit der Konsumenten führen, die *expressive dimension*, und Attribute welche zu sinkender Gesamtzufriedenheit der Konsumenten führen, die *instrumental dimension*. So ist Kundenzufriedenheit kein unipolares Konstrukt, sondern bipolar. Zufriedenheit und Unzufriedenheit können sich nicht gegenseitig kompensieren, da sie unabhängige Konstrukte sind.

Es wird eine neuronale Netzwerkmethod zur Klassifizierung der Zufriedenheitslinks zwischen den Attributen und der Gesamtzufriedenheit verwendet. Der Link zwischen den Attributen und der Gesamtzufriedenheit, kurz A-OSL (*attributes-overall satisfaction link*), wird explorativ mit der Kohonen Self-Organizing Map (SOM), welche in genanntem Artikel auch detailliert beschrieben wird, untersucht. Dabei werden Prototypen der Links aufgedeckt.

Zusätzlich werden die Daten in Bezug auf diese Prototypen auf Basis des euklidischen Distanzkriteriums klassifiziert. Kohonen SOM erschafft einen topologischen Grid, auf welchem die Prototypen aufeinander bezogen werden. Fünf A-OS Links werden vorgeschlagen, lineare A-OSL, Attribute mit steigenden Erträgen in Bezug auf TCOS (Kundengesamtzufriedenheit), Attribute mit sinkenden Erträgen in Bezug auf TCOS, Attribute mit steigenden Erträgen in Bezug auf TCOS und Kundengesamtzufriedenheit, und Attribute, welche einen *assimilation-contrast* Effekt aufweisen. Diagramm 9 beschreibt Kahneman und Tversky's *prospect theory*, welche eine sigmoide Funktion empfiehlt. Sie geht davon aus, dass sich ein Verlust stärker auf die Präferenz auswirkt, als ein Gewinn. Zusätzlich zeigt die Performance sinkende Erträge der Kundenpräferenzen. Auf die Kundenzufriedenheit übersetzt bedeutet dies, dass sich negative Performance mehr auswirkt als positive, und die Performance sinkende Erträge hinsichtlich der Gesamtzufriedenheit aufweist. So muss die A/OS einen S-förmigen Verlauf annehmen.

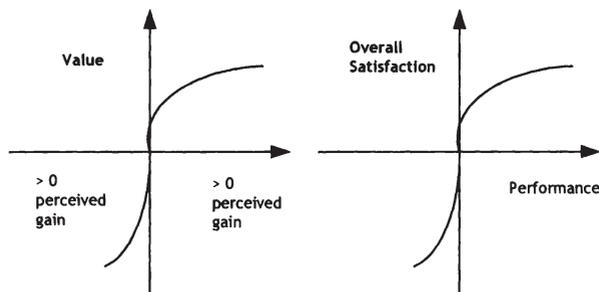


Diagramm 9: Prospect Theory (Audrian-Pontevia, 2006)

Die vier unterschiedlichen Links mit den einhergehenden Managementempfehlungen werden in Tabelle 7 aufgelistet.

Die Vorteile der von Audrian-Pontevia (2006) vorgestellten Kohonen SOM Methode liegen in der Bestimmung der Anzahl an Prototypen, hier die Funktionen für die A-OSL der Daten. Zusätzlich werden die Daten in Relation zu den Prototypen klassifiziert. Schließlich kann Kohonen SOM eine große Datenmenge verarbeiten und schiefلاغige Daten, wie beispielsweise Zufriedenheitsdaten, verwerten. Somit stellt die Robustheit gegenüber den Nicht-Normalitätsannahmen einen zusätzlichen Vorteil dar. Die optimale Lösung ist jene, welche die Intraklassendistanzen, verglichen mit der durchschnittlichen Interklassendistanz, gering hält, hier eine 4x4 Map, wie in Diagramm 10 dargestellt. Die Visualisierung des topologischen Grid hilft die Funktionen zu

verstehen und zu analysieren. Lineare als auch nichtlineare Zusammenhänge werden in der linken Graphik aufgefunden. Ein Beispiel gibt die rechte Graphik in Diagramm 10. Die horizontale Achse gibt den durchschnittlichen Attributzufriedenheitslevel an, die vertikale Achse den Gesamtzufriedenheitslevel. Die ersten vier Knickpunkte der Kurve repräsentieren den Clusterprototypen. Um jeden Prototyp geben zwei Kurven die $\pm 1,5$ Standardabweichung an, welche eine visuelle Information zur Klassifikationsqualität geben. Sind die Kurven in der Nähe des Prototyps, ist die Klassifizierung homogen. Die Clustergröße, also die Anzahl an klassifizierten Attributen, wird durch die Kreisgröße dargestellt. Wie in Tabelle 7 ist Prototyp (4,3) aus Diagramm 10 (rechte Graphik) als *satisfaction maintainer* zu klassifizieren.

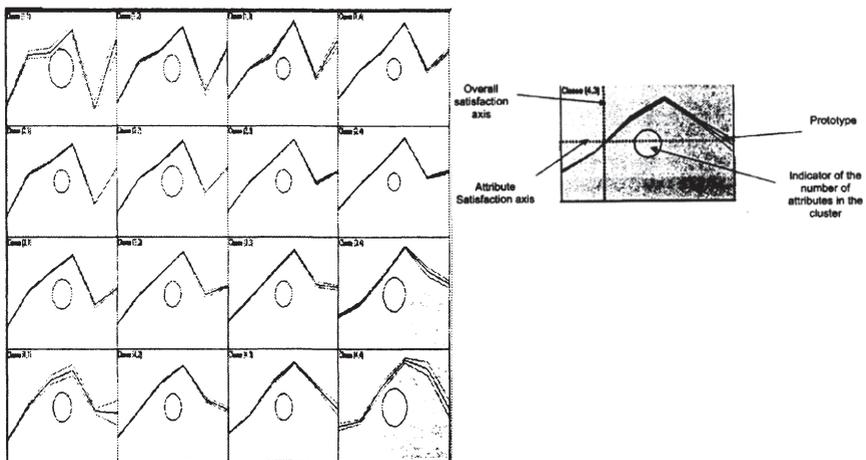


Diagramm 10: Kohonen Self-Organizing Map mit Satisfaction Maintainer (Audrian-Pontevia, 2006)

Die Existenz linearer und nicht-linearer/asymmetrischer A-OSL kann gezeigt werden. Lineare oder quasi-lineare Prototypen, welche für *one-dimensional* Attribute stehen (z.B. Cluster 3,3 oder 2,4), Prototypen mit steigenden Erträgen in Bezug auf die Gesamtzufriedenheit, welche für *attractive* Attribute stehen (z.B. Cluster 3,4), Prototypen mit sinkenden Erträgen in Bezug auf die Gesamtzufriedenheit, welche für *must-be* Attribute stehen (z.B. Cluster 4,1), Prototypen mit steigenden Erträgen in Bezug auf die Gesamtzufriedenheit und die Unzufriedenheit, welche für *satisfaction maintainers* stehen (z.B. Cluster 4,4), als auch Attribute mit steigenden Erträgen in Bezug auf die Gesamtzufriedenheit mit einem Schwelleneffekt (z.B. Cluster 1,1), können gefunden werden. Letzterer zeigt eine neutrale Zone um den Mittelpunkt, die *indifference*

zone, innerhalb welcher die Performance keinen signifikanten Einfluss auf die Gesamtzufriedenheit hat, außerhalb dieser jedoch einen starken Einfluss ausübt. Der Bereich liegt hier etwa bei einer Standardabweichung von 1, innerhalb welcher die Kundengesamtzufriedenheit nicht signifikant beeinflusst wird. Bei einer positiv oder negativ größeren Standardabweichung der Performance über 1 ist der Einfluss auf die Gesamtzufriedenheit bzw. –unzufriedenheit sehr groß. Dies steht für den *assimilation-contrast*-Effekt der Performance auf die Gesamtzufriedenheit. *Must-be* oder *satisfaction maintainers* beeinflussen am meisten die Kundengesamtzufriedenheit. Andere wie *attractive* oder *assimilation-contrast* entfalten ihr Potential erst in höheren Levels der Kundengesamtzufriedenheit.

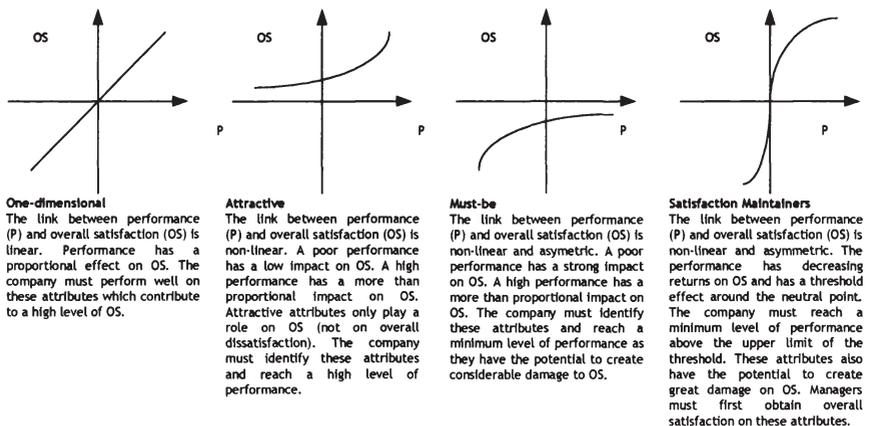


Tabelle 7: Attribute – Overall Satisfaction Links (Audrian-Pontevia, 2006)

Auch die zeitabhängige Verwendung wird im Zuge dieses Verfahrens angesprochen. Zusätzlich zur Aufdeckung benachbarter Eigenschaften, sollte die Attributs/Gesamtzufriedenheitsfunktion auch über die Zeit hinweg untersucht werden. Da es möglich ist neue Vektoren auf ein vordefiniertes Grid zu projizieren, kann Benchmarking über die Zeit oder zwischen Produkten/ Serviceleistungen ermöglicht werden.

3.2. Modellvergleich

3.2.1. Importance Grid vs. Penalty-Reward-Contrast Analysis

Im Folgenden wird genauer auf den Unterschied zwischen direkt und indirekt gemessener Wichtigkeit und den sich daraus ergebenden differenzierten Interpretationen eingegangen. Matzler und Sauerwein (2002) durchleuchten die

Eigenheiten der einzelnen Attribute. Diskutiert werden die Annahmen des zweidimensionalen *Vavra's Importance-grids*, basierend auf selbstbeurteilter Wichtigkeit und abgeleiteter Wichtigkeit unter Verwendung einer Regressionsanalyse, und die der *penalty-reward contrast* (PRC) Analyse von Brandt. Letztere unterteilt die Attribute mit Hilfe einer Dummy-Regression in drei Faktoren, die *penalty-factors* (Minimalansprüche), *reward-factors* (wertsteigernde Merkmale) und *hybrid factors* (Einfluss auf Zufriedenheit und Unzufriedenheit).

Frühe Studien unterschieden lediglich zwischen *satisfiers* und *dissatisfiers*. Spätere beziehen einen dritten Faktor ein, welcher zu Zufriedenheit und Unzufriedenheit führt. Diese Erweiterung wird unterstützt durch Methoden wie *critical incident technique* (CIT), Inhaltsanalyse von Beschwerden und Komplimenten, Rangordnung der Serviceattribute für gutes und schlechtes Service und auch Regressionsanalyse. Auch der Zusammenhang der drei Faktoren zwischen Zufriedenheit und Wichtigkeit wurde untersucht. *Basic*-Faktoren führen nicht zu Zufriedenheit, wenn die Performance hoch ist. Ihr Zufriedenheitseinfluss, also ihre relative Wichtigkeit, ist niedrig. Ist die Performance niedriger als erwartet, führt dies zur Unzufriedenheit und die relative Wichtigkeit ist hoch. *Excitement*-Faktoren führen, wenn vorhanden, zu *delight* und üben einen starken Einfluss auf die Zufriedenheit auf. Werden diese nicht erfüllt, führt dies nicht zur Unzufriedenheit. Somit ändert sich die relative Wichtigkeit, wenn sich die Performance ändert. Die Wichtigkeit ist folgend vom derzeitigen Zufriedenheitslevel abhängig. Wird eine implizite Wichtigkeitsmessung verwendet, beispielsweise die Attributskorrelation mit einem externen Kriterium wie Gesamtzufriedenheit, so hängt diese auch vom derzeitigen Zufriedenheitslevel ab. So sind die impliziten Wichtigkeiten, welche über die relativen Gewichte bestimmt wurden, auch unterschiedlich, verglichen mit den selbstbeurteilten. Selbstbeurteilte Attributswichtigkeiten, also expliziten Wichtigkeiten, sind keine Funktion der Attributzufriedenheit. Sie reflektieren nicht die Zufriedenheits-Wichtigkeits-Beziehung. Dieses Argument festigen sie durch den Umstand, dass die implizit abgeleiteten Attributswichtigkeiten eine Funktion der Attributzufriedenheit sind und es nicht-lineare Beziehungen zwischen der Attributlevelzufriedenheit und der Gesamtzufriedenheit gibt. Die Unterteilung der x- und y-Achse nach Mittelwert bzw. Median führt zu Tabelle 8. Explizite und implizite Wichtigkeiten unterscheiden sich.

Implicit importance (derived importance)	<i>High</i>	High implicit importance/ low explicit importance = excitement factors	High implicit importance/ high explicit importance = performance factors (important)
	<i>Low</i>	Low implicit importance/ low explicit importance = performance factors (unimportant)	Low implicit importance/ high explicit importance = basic factors
		<i>Low</i>	<i>High</i>
Explicit importance (customer's self-stated importance)			

Tabelle 8: Importance Grid (Matzler und Sauerwein, 2002)

Die Ergebnisse des *Importance grid* werden mit den Ergebnissen der PRC Analyse verglichen. Für die PRC werden die ‚sehr zufrieden‘-Bewertungen verwendet um Dummy-Variablen für die *excitement*-Faktoren zu erhalten (Wert 0). ‚Etwas unzufrieden‘- und ‚sehr unzufrieden‘-Bewertungen führen zu Dummy-Variablen für die *basic*-Faktoren (Wert 1). ‚Etwas zufrieden‘- und ‚weder zufrieden noch unzufrieden‘ stehen für Indifferenz wo die Erwartungen erfüllt wurden und bilden die Referenzgruppe. Die unabhängigen Dummy-Variablen, als *rewards* und *penalties* bezeichnet, werden in der multiplen Regression auf die 5-Punkt-Likert-Gesamtzufriedenheitsskala als abhängige Variable regressiert. Die Konstante der Regressionsgleichung entspricht dem Durchschnitt der Referenzgruppen der Gesamtzufriedenheit. *Penalties* steht für jenen Betrag, welcher von der Konstante bei niedriger Zufriedenheit mit einem Attribut abgezogen werden muss, *rewards* für jenen, der bei hoher Zufriedenheit hinzugezählt wird. Das Ergebnis wird in Diagramm 11 gezeigt. Überwiegt *penalty* gegenüber *rewards*, so ist das Attribut ein *basic*-Faktor. Überwiegt *rewards* gegenüber *penalty*, so ist das Attribut ein *excitement*-Faktor. Sind beide gleich, so spricht man von einem *performance*-Faktor bzw. *hybrid*. Zufriedenheit entsteht wenn die Performance hoch ist und Unzufriedenheit wenn diese niedrig ist.

Um die beiden Methoden zu beurteilen, werden diese miteinander verglichen. *Convergent validity* kann nicht bestätigt werden, da die beiden Methoden eine unterschiedliche Einteilung der Faktoren vorschlagen. Kritik wird an der Methode mit den expliziten Wichtigkeiten geübt. Die direkten Attributwichtigkeitsmessungen scheinen mehrdeutig und nicht reliabel. Außerdem ist die Attributwichtigkeit eine Funktion der Performance. Weiters sind *basic*-Faktoren nur wichtig, wenn sie nicht zufriedenstellend erfüllt werden. Die Konsumenten beziehen in die explizite Wichtigkeitsbewertung den aktuellen Level der Zufriedenheit nicht mit ein. Die Regressionslösung errechnet diese auf Basis der aktuellen Zufriedenheit. Weiters wird die Wichtigkeits-Zufriedenheitsbeziehung

nicht berücksichtigt. So muss auch die ursprüngliche IPA-Analyse von Martilla und James (1977) hinterfragt werden, da die Zufriedenheiten und Wichtigkeiten korreliert sind. Explizite Wichtigkeitsgewichte können auch sozial akzeptable, politisch korrekte oder strategisch bedingte Antworten repräsentieren. Weiters ist die Klassifikation über die Mittelwerte der Koordinaten der Matrix mehr oder weniger willkürlich. Der *Importance-grid* kann lediglich Zufriedenheitsfaktoren identifizieren, PRC hingegen misst zusätzlich, durch die Verwendung von Regressionskoeffizienten für niedrige und hohe Performance, die asymmetrische Beziehung zwischen Attributlevel- und Gesamtzufriedenheit. PRC kann wiederum aufgrund der Dichotomisierung nur wenig Varianz erklären. Es wird also festgestellt, dass die Wichtigkeitsgewichte abhängig von der Messung, also implizit vs. explizit, sind. Abschließend werden Handlungsempfehlungen in Tabelle 9 dargestellt. Es ist erkennbar, wie sich die Wichtigkeit der Basis- und Excitementfaktoren ändert, wenn sich die Attributzufriedenheit ändert. Die Wichtigkeit der *basic*-Faktoren wird unterschätzt, wenn die Performance hoch ist und überschätzt, wenn diese niedrig ist. Bei *excitement*-Faktoren wird der Einfluss unterschätzt, wenn die Performance niedrig ist. Die Veränderung der Attributzufriedenheit löst eine Veränderung der Attributswichtigkeit aus. Deshalb ist es wichtig die relative Wichtigkeit jedes Attributes für hohe und niedrige Performance zu messen. Diese können jedoch nur richtig bestimmt werden, wenn Nichtlinearitäten berücksichtigt werden.

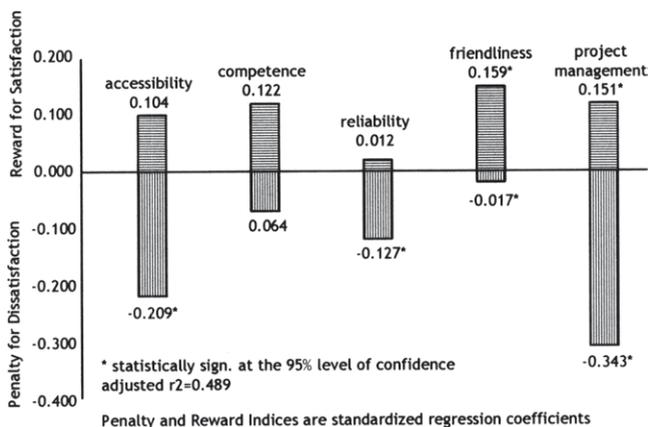


Diagramm 11: Penalty-reward contrast-Analyse (Matzler und Sauerwein, 2002)

Fuchs und Weiermair (2004) vergleichen ebenfalls die zwei Methoden, Vavra's zweidimensionalen *Importance-grid* basierend auf direkt abgefragter Wichtigkeit

und über partielle Regression hergeleitete Wichtigkeit, zum anderen die PRC-Analyse, welche lediglich auf Zufriedenheitsbeurteilungen basiert und ein Dummy-Regressionsmodell benutzt. Vavra's Ansatz geht von der Überlegung aus, dass die Konsumenten fähig sind, zwischen expliziter und impliziter Wichtigkeitsdimensionen eines Produktes/Service zu unterscheiden. Dies führt zur Identifizierung von drei unterschiedlichen Zufriedenheitsdeterminanten, *satisfiers*, *performance*-Faktoren und *dissatisfiers*.

Attribute	<i>High</i>	Low importance	High importance ↑
performance	<i>Low</i>	High importance ↓	Low importance ↑
		<i>Basic factors</i>	<i>Excitement factors</i>

Tabelle 9: Importance-grid Empfehlungen (Matzler und Sauerwein, 2002)

Der Artikel beginnt mit einer einleitenden Diskussion zur Messung von Zufriedenheit. Zufriedenheit kann als Konsumentenbeurteilung einer Konsumationserfahrung, basierend auf der Beziehung zwischen der subjektiven Wahrnehmung des Konsumenten und den objektiven Attributen des Produktes gesehen werden. Selbiges gilt in Bezug auf das Service, wobei die Interaktion zwischen den gemachten Erfahrungen und den Erwartungen berücksichtigt wird. Laut der Amerikanischen Schule, Parasuraman, Zeithaml und Berry, stellt sie eine Reflexion positiver oder negativer *Gaps* zwischen Erwartung und Wahrnehmung dar. Die zentrale Idee des *Expectation-Performance* Ansatzes stammt aus der Differenz zwischen empfundener Performance und empfundener Wichtigkeit. Die Skandinavische Schule der Servicequalität um Grönroos ignoriert die Erwartungen und verwendet einen reinen Performanceansatz, da es unmöglich ist, hohe Zufriedenheitslevels als Resultat niedriger Erwartungen oder hoher Qualität zu interpretieren. Den beiden Sichtweisen stehen jedoch etliche Kritikpunkte gegenüber. So wird die Existenz der Uneinstimmigkeit unter einer korrekten Definition der Erwartungen unterstellt. Diese kann als Prognose zukünftiger Performance, Wünsche, Ziele oder auch als Normwerte basierend auf vorangegangenen Erfahrungen gesehen werden. Probleme gibt es beispielsweise bei der Erwartungsbeurteilung, wenn das Service nur selten in Anspruch genommen wird, sich Erwartungen wenn auch nur kurzfristig aufgrund zusätzlicher Information ändern, aufgrund der Neuartigkeit keine genauen Erwartungen bestehen, oder kulturelle Unterschiede in Einstellung und Verhalten die Erwartungen und die Wahrnehmung beeinflussen. Somit ist es schwer zwischen Erwartung und wahrgenommener Performance zu unterscheiden. Sie schlussfolgern, dass trotz zahlreicher Verwendung des *Expectation-Performance* Ansatzes der reine Performanceansatz eine gute Alternative darstellt. Vavra's *Importance-grid* schreiben sie eher dem

Erwartungs-Performance Ansatz der Kundenzufriedenheit zu, die PRC-Analyse eher dem reinen Performanceansatz, da letzterer rein Zufriedenheitsvariablen verwendet.

Unterschiede werden zwischen beiden Methoden aufgefunden, als auch innerhalb dieser, wenn die Befragten in Segmente unterteilt werden. Die Vorteile des *Importance-grid* liegen in der Untersuchung des Zusammenhanges zwischen Zufriedenheit und Wichtigkeit und der Möglichkeit, eine Vielzahl an Attributen einfließen zu lassen. Nachteile werden in der quasi beliebigen Unterteilung der Faktoren anhand der Mittelwerte und der direkten Messung der Wichtigkeit und deren Veränderung im Zuge der Konsumation, aufgrund der Beeinflussung durch die Zufriedenheit, gesehen. Zusätzlich steckt die Annahme im Modell, dass die Befragten fähig sind, zwischen den beiden Wichtigkeitsarten zu unterscheiden. Die Korrelationsanalyse zur Berechnung der impliziten Wichtigkeit wird ebenfalls bemängelt, da sie lediglich den Zusammenhang und nicht die Regressionsanalyse wie im Ansatz von Brandt verwendet. Zusammenfassend wird dem reinen Performanceansatz von Brandt für kleinere Sets an unabhängigen Variablen der Vorzug gegeben.

Füller et al. (2006) stellen die Attributswichtigkeit ebenfalls als Funktion der Attributperformance dar. Zufriedenheit wird nämlich durch das Erreichen einer wichtigen Leistung erzielt. Sie erwähnen die *Prospect Theory* als Begründung für den asymmetrischen Einfluss der Attributlevelperformance auf die Gesamtzufriedenheit und das Wiederkaufverhalten. Diese verleiht niedriger Attributperformance einen größeren Einfluss auf die Gesamtzufriedenheit, verglichen mit höherer. Die Asymmetrie ist attributabhängig. Mit Hilfe der multiplen Dummy-Variablen Regression können die Faktoren *dissatisfier*, *satisfier* und *hybrids* unterschieden werden. Im Zuge der Dummy-Variablenerstellung wird ein Set an Variablen für *satisfier* geschaffen, wobei Werte im niedrigsten Quartil mit 1 codiert werden. Ein weiteres Set wird für *dissatisfier* geschaffen, wobei Werte im höchsten Quartil mit 1 codiert werden. Tabelle 10 zeigt das Ergebnis. ‚Waiting time‘ stellt einen *dissatisfier* dar. In niedrigen Zufriedenheitsbereichen ist der Regressionskoeffizient signifikant und in höheren nicht signifikant. ‚Restaurants‘ und ‚bars‘ sind *hybrids*. Die Wichtigkeit hängt nicht von der Performance ab. ‚Slopes‘, ‚fun‘, ‚entertainment‘ und ‚employees‘ zeigen einen schwach stärkeren Einfluss wenn die Zufriedenheit gering ist. Eine Veränderung der Attributperformance (Zufriedenheit) kann mit der Veränderung der Attributswichtigkeit verknüpft werden. Deshalb muss der relative Einfluss jedes Attributes bei hoher und bei niedriger Performance

geschätzt werden. Die bis dato vorgestellten Modelle mit mehreren Attributen sollten folglich überdacht werden und die Asymmetrie berücksichtigen.

Dependent variable: overall satisfaction	Dummy-variable regression coefficients	
	High satisfaction	Low satisfaction
Slopes	.09 ^a	-.12 ^a
Restaurants and bars	.14 ^a	-.14 ^a
Fun & entertainment	.08 ^a	-.13 ^a
Waiting times	.01ns	-.08 ^a
Employees	.09 ^a	-.13 ^a

R² .394; ^ap<.001, ns = not significant, F-Value 163.134 (Sig. = .000).

Tabelle 10: Asymmetric Impact of Satisfaction Drivers (Füller et al., 2006)

3.2.2. Importance-Performance Matrix vs. Dummy-Regression

Matzler et al. (2004) vergleichen die IPA mit der Dummy-Regression. Hinter der IPA stehen die Annahmen, dass Attributperformance und –wichtigkeit unabhängig voneinander sind und die Beziehung zwischen Attributqualitätsperformance und Gesamtleistung linear und symmetrisch ist. Dies erklärt sich dadurch, dass die Wichtigkeit eines *basic*- oder *excitement*-Faktors von der Performance abhängig ist. *Basic*-Faktoren sind entscheidend, wenn die Performance niedrig ist und unwichtig, wenn diese hoch ist. Umgekehrtes gilt für *excitement*-Faktoren. Die Drei-Faktoren-Theorie der Kundenzufriedenheit widerspricht der Ansicht, dass die relative Wichtigkeit von Serviceattributen durch eine Punktschätzung repräsentiert werden kann. Vielmehr ist sie eine Funktion der Zufriedenheit. Aus dem Kano-Modell folgen weitere Annahmen, nämlich die Nichtlinearität und Asymmetrie der Beziehung zwischen Performance und Wichtigkeit von *basic*- und *excitement*-Faktoren und die Attributwichtigkeit als Funktion der Performance. Dies hat unterschiedliche Auswirkungen wenn die Attributperformance aus dem *disconfirmation* Paradigma von Oliver betrachtet wird. Hierbei wird angemerkt, dass die Zufriedenheit ein kognitiver Vergleich zwischen empfundener Performance mit vorangehenden Erwartungen ist. Die empfundene Performance kann größer als die Erwartung sein und führt zu *positive confirmation* (Zufriedenheit), oder sie kann niedriger sein und führt zu *negative disconfirmation* (Unzufriedenheit). Ist die Performance wie erwartet, resultiert daraus mittlere Zufriedenheit oder Indifferenz.

Performance wird in der IPA direkt abgefragt. Die Attributwichtigkeit resultiert aus der multiplen Regression zwischen der Gesamtzufriedenheit als Abhängige

und der Attributperformance als Unabhängige. Für die Performancebewertungen im Zuge der Dummy-Variablen Regression werden ‚niedrige Performance‘ (0,1), ‚hohe Performance‘ (1,0), und ‚durchschnittliche Performance‘ (0,0) in Dummy-Variablen recodiert. So erhält man für jede Variable zwei Regressionskoeffizienten. Einer steht für den Einfluss, wenn die Performance niedrig ist, der andere wenn diese hoch ist. So wird der asymmetrische Einfluss der Attributlevelperformance auf die Gesamtpformance gemessen. Diagramm 12 zeigt die Abhängigkeit der Wichtigkeit von der Performance.

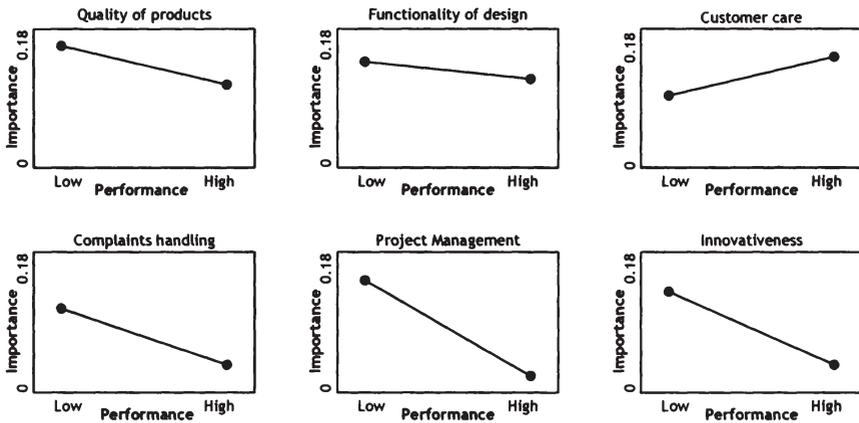


Diagramm 12: Wichtigkeit als abhängige Variable von Performance (Matzler, 2004)

Aus Diagramm 13 werden Fehlinterpretationen der IPA abgeleitet. Dafür wurde die Stichprobe in zwei Teile geteilt und für beide Teile getrennt die IPA visualisiert. Als Trennkriterium wurde die Zufriedenheitsskala herangezogen. 7 bis 10 bildet die zufriedene Gruppe (rechte Graphik), 1 bis 6 die unzufriedene Gruppe (linke Graphik).

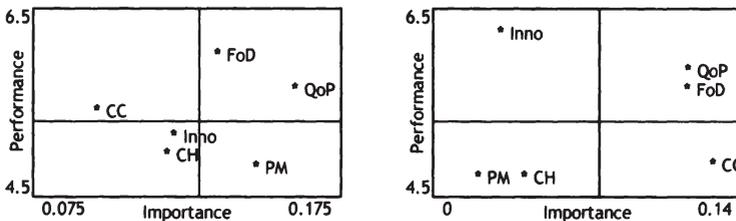


Diagramm 13: Gruppenvergleich der IPA matrix (Matzler, 2004)

Für beide Gruppen folgen unterschiedliche Handlungsempfehlungen. Die IPA ist also als Entscheidungsunterstützung zwecks Ressourcenverteilung nicht brauchbar, da Attributswichtigkeit und -performance nicht unabhängig sind.

3.2.3. Zufriedenheitskonstrukte

Vargo et al. (2007) liefern einen Überblick über die Einteilung in *satisfiers*, *dissatisfiers*, *criticals* und *neutrals*. Unterschiedliche Komponenten haben unterschiedliche Fähigkeiten Zufriedenheit bzw. Unzufriedenheit zu beeinflussen. Tabelle 11 gibt einen Überblick zur Aufteilung der Auslöser von Zufriedenheit und Unzufriedenheit vorangegangener Theorien. Die Zwei-Faktoren Theorie von Herzberg, auch Motivator-Hygiene Modell genannt, teilt sie in Motivatoren, welche zu Zufriedenheit führen, aber bei Abwesenheit nicht zu Unzufriedenheit, und in Hygienefaktoren, welche keine Zufriedenheit auslösen, aber bei Abwesenheit Unzufriedenheit bewirken. Die Motivatoren und Hygienefaktoren werden mit den *expressive*- und *instrumental*-Faktoren verbunden. *Expressive*-Faktoren beschreiben psychologische Aspekte, *instrumental*-Faktoren physische Aspekte.

Herzberg et al. (1959) Soliman (1970) Zahng & Dran (2000)	Motivators	Hygienes	-	-
Swan & Combs (1976)	Expressive factors	Instrumental factors	-	-
Kano et al. (1984) Matzler & Hinterhuber (1998)	Attractive	Must-be	One-dimensional	Indifferent
Oliver (1995)	Monovalent satisfiers	Monovalent dissatisfiers	Bivalent satisfiers	Null relationships
Cadotte & Turgeon (1988)	Satisfiers	Dissatisfiers	Criticals	Neutrals

Tabelle 11: Einteilung der Auslöser von Satisfaction und Dissatisfaction (Vargo et al., 2007)

Diagramm 14 visualisiert die Idee von Kano mit den Elementen, *attractive*, *must-be*, *one-dimensional*, *indifferent* und *reverse quality*.

Diagramm 15 beschreibt eine Einteilung, welche die Beziehung zwischen Bedürfnisbefriedigung und Zufriedenheit beschreibt. *Monovalent dissatisfier* (*dissatisfiers*) liefern den größten Teil der Unzufriedenheit, *monovalent satisfiers* (*satisfiers*) liefern den größten Teil der Zufriedenheit, *bivalent satisfiers* (*criticals*) beeinflussen sowohl Zufriedenheit als auch Unzufriedenheit. Es wird

jedoch angemerkt, dass *monovalent satisfiers* und *dissatisfiers* auch einen bescheidenen Teil zum gegensätzlichen Zufriedenheitsbereich beitragen.

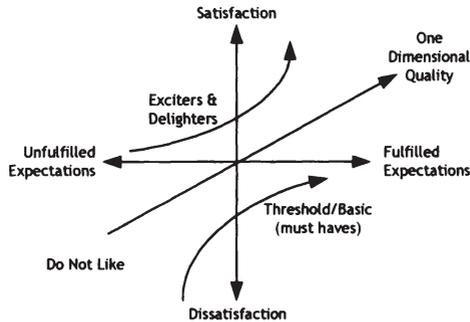


Diagramm 14: Kano Modell (Vargo et al., 2007)

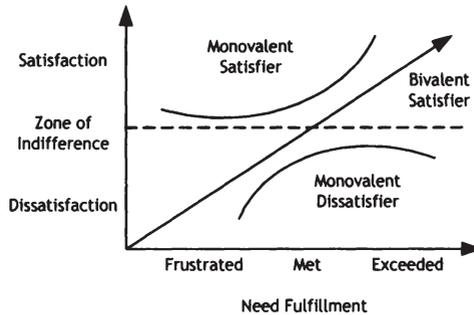


Diagramm 15: Monovalent vs. bivalent (Vargo et al., 2007)

Diagramm 16 stellt die hypothetische Verteilung der empfundenen Performance dar. *Dissatisfier* sind schief in Richtung der negativen Performanceempfindung. *Satisfier* sind schief in Richtung der positiven Empfindung. *Neutrals* und *criticals* zeigen eine ausgewogenere Verteilung. Den Unterschied machen die Platzierung und die Dominanz der Indifferenzzone, welche die neutralen Beurteilungen repräsentiert, aus.

Auch bezüglich der Wichtigkeit der einzelnen Komponenten gibt es Unterschiede. *Dissatisfiers* haben meist eine extremere Wirkung auf die Empfindung, verglichen mit *satisfiers*. Anhand der *prospect theory*, kann der Nutzenwert asymmetrisch repräsentiert werden. Die Wertefunktion kann als Gewinn (*satisfier*) oder Verlust (*dissatisfier*) zu einem Referenzpunkt gesehen werden. Dies führt zu einer S-kurvigen Funktion relativ zu Gewinnen und Verlusten. Für Gewinne ist diese meist konkav (risikoavers) und für Verluste konvex (risikoaffin). Die *value*-Funktion ist aufgrund der Risikoaversion steiler

für Verluste als für Gewinne. Der Grenzwert der Gewinne und Verluste sinkt mit deren Größe aufgrund der *diminishing sensitivity*. Aufgrund dieser Sensitivität sind Eingriffe um Verluste zu minimieren, Elimination von *dissatisfiers*, wichtiger wenn das Attributlevel niedriger als, oder in der Nähe des Referenzpunktes liegt, als die Erhöhung von Gewinnen, *satisfiers*, wenn das Level bereits weit über dem Referenzpunkt liegt. *Criticals* wirken meist stärker als *satisfiers* oder *dissatisfiers*, da sie die Kerneigenschaften des Produktes repräsentieren.

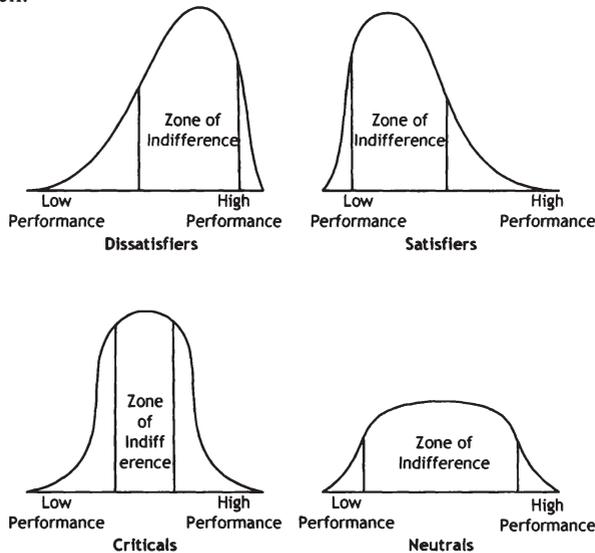


Diagramm 16: Zone of Indifference (Vargo et al., 2007)

Vargo et al. (2006) sehen die zeitliche Veränderung von *satisfiers* zu *criticals* und dann zu *dissatisfiers*. Das *augmented product* wird *expected*, *generic* und schließlich *dissatisfier*.

Sie beschreiben drei unterschiedliche Ansätze. Der erste Ansatz, das *curvilinear (dis)satisfaction*-Modell, unterstreicht die Existenz von *satisfiers*, *dissatisfiers*, *criticals* und *neutrals*. *Monovalent dissatisfiers* und *monovalent satisfiers* sind kurvilinear. *Bivalent satisfiers/criticals* zeigen eine monoton wachsende Beziehung. Eine Null- oder Zufallsbeziehung bedeutet, dass die Bedürfniserfüllung nicht mit der Zufriedenheit zusammenhängt. Der zweite Ansatz im *context-specific (dis)satisfaction framework*, beschreibt die Zufriedenheitsmessinstrumente in Abhängigkeit davon, ob diese *satisfier*, *dissatisfier*, *criticals* oder *neutrals* sind. Werden eher *dissatisfier* untersucht sind Messungen wie

‚relieved‘ oder ‚skeptical‘ besser geeignet. Um *satisfiers* zu messen eignen sich beispielsweise ‚excited‘ und ‚pleased‘. Der dritte Ansatz, der *social judgment theory approach*, teilt nach Einstellungsbereichen auf. *Latitude of acceptance* beschreibt die am meisten akzeptablen, *latitude of rejection* die störenden, und *latitude of non-commitment* die weder akzeptabel noch als störend erachteten. Die asymmetrische Natur dieser Spielräume stimmt mit der Einteilung in *satisfiers*, *dissatisfiers*, *criticals* und *neutrals* überein. *Satisfiers* werden über *latitude of acceptance*, *dissatisfiers* über *latitude of rejection* und *neutrals* über *latitude of non-commitment* gemessen.

Weiters werden drei Messmöglichkeiten der *satisfiers*, *dissatisfiers*, *criticals* und *neutrals* beschrieben. Die *critical incident technique* (CIT) identifiziert rein zufriedene oder unzufriedene Sachverhalte. Diese werden in eine 2x2-Matrix nach hoher und niedriger Zufriedenheit und Unzufriedenheit, oder verwandte Messungen wie Komplimente oder Beschwerden, eingeteilt. Ereignisse die nur Zufriedenheit auslösen sind *satisfiers*, jene die nur Unzufriedenheit auslösen sind *dissatisfiers*. Die Schwäche liegt in der reinen Messung der Extreme, in dem Umstand, dass das Ereignis bereits stattgefunden haben muss, im zeitlichen Aufwand und in der Objektivität und Reliabilität die Anekdoten zu analysieren. Die zweite Messmöglichkeit, die *functional and dysfunctional technique* von Kano stellt zwei unterschiedliche Fragen. Bei *functional* wird das Vorhandensein eines hohen Levels angenommen, bei *dysfunctional* ein niedriges Level. Geantwortet wird mit ‚like‘, ‚must-be‘, ‚neutral‘, ‚i can live with it that way‘ und ‚dislike‘. In der resultierenden 5x5-Matrix repräsentiert eine Dimension die funktionale und eine die dysfunktionalen Bedingungen. Durch Verbindung der beiden Fragen werden die Items in *attractive*, *must-be*, *one-dimensional*, *indifferent*, *reverse* oder *questionable/skeptical* kategorisiert. Die Häufigkeit entscheidet über die Einteilung. Die anschließende Berechnung des *consumer satisfaction* (CS) Koeffizienten, gibt den Grad der möglichen Steigung von Zufriedenheit oder Senkung von Unzufriedenheit an. Kritik wird an der Skalierung geübt, welche bestenfalls ordinal ist, und später in eine nominale übergeführt wird. Ein intervallskalierter Index wäre besser. Die dritte Technik, die *need gratification technique*, fragt danach, ob die Performance die Bedürfnisse erfüllt hat, zu kurz gekommen ist, oder diese übertroffen hat. Dann wird die Beziehung zwischen den Antworten auf der Bedürfnis-Gratifizierungsskala und der Gesamtzufriedenheit betrachtet. Die Beziehung sollte *curvilinear* für *monovalent dissatisfiers* und *monovalent satisfiers* sein, mit einigen *bivalent satisfiers/criticals*, welche eine monoton steigende Beziehung zeigen. Eine Null- oder Zufallsbeziehung zeigt, dass die Bedürfnisbefriedigung nicht mit der Zufriedenheit zusammenhängt. Die *three-need category scale* hat

zwar das Potential eines quantitativen Index, wurde jedoch noch nicht weiterentwickelt.

3.3. Segmentierung

3.3.1. Spezifizierung eines erweiterten Kano-Modells

Buhl et al. (2007) erläutern unterschiedliche Perspektiven des Kano-Modells. In der Kundenzufriedenheitsforschung wird weitgehend der *Confirmation/Disconfirmation (C/D)*-Ansatz verwendet, welcher Zufriedenheit als Einstellung aus einem Soll-Ist-Vergleich definiert. Die Ist-Leistung ist die wahrgenommene Qualität, die Soll-Komponente ist ein subjektiver Vergleich in Form einer Erwartungshaltung. Entsprechen sich beide, so wird das Konfirmationsniveau der Zufriedenheit erreicht, übertreffen die Leistungen die Erwartungen spricht man von positiver Diskonfirmation bzw. Zufriedenheit, umgekehrt von negativer Diskonfirmation bzw. Unzufriedenheit. In der letzten Formulierung wird Diskonfirmation als generelle Nicht-Übereinstimmung zwischen Erwartung und Erfüllung gesehen. Konfirmation entspricht somit der Übereinstimmung. Diese Formulierung geht jedoch nicht mit der Sichtweise des C/D-Paradigmas konform. Der C/D-Ansatz misst verstärkt die kognitive und weniger die affektive Evaluierung der Performance und ist somit eher für funktionale Eigenschaften geeignet, wie auch das ursprüngliche Kano-Modell. Übertragen auf die Kundenzufriedenheit entspricht dies der Erfüllungserwartung.

Auch die zeitliche Komponente, in Diagramm 17 dargestellt, wird angesprochen. Begeisterungsfaktoren (BeF) werden zu Leistungsfaktoren (LF) und LF zu Basisfaktoren (BaF), da aktuell nicht erwartete Leistungen in der Zukunft bereits erwartet werden. Dies führt zu sinkender Zufriedenheit über die Zeit, sofern keine weiteren außergewöhnlichen Leistungen gefunden werden.

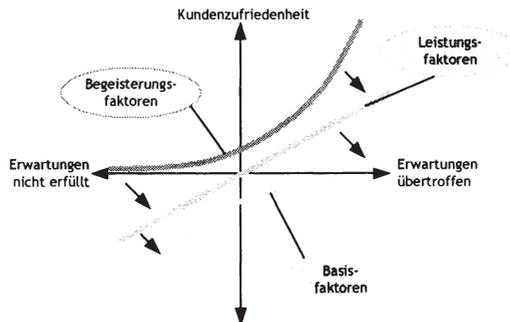


Diagramm 17: Kano-Modell übertragen auf die Kundenzufriedenheit (Buhl et al., 2007)

Buhl et al. (2007) merken an, dass nicht alle Kunden gleich auf bestimmte Leistungen hinsichtlich Zufriedenheit reagieren. Diese Heterogenität muss mit mehreren homogenen Kundensegmenten aufgefangen werden, da unterschiedliche Segmente die gleiche Leistung in unterschiedliche Kano-Faktoren einteilen können. Verglichen wird jede Leistung mit den Erwartungen eines Segmentes in einer bestimmten Periode. Dieser Vergleich steht für die Erfüllungserfüllung. Es wird davon ausgegangen, dass das Leistungsniveau über die Zeit gleich bleibt, die Erwartungen jedoch monoton ansteigen. Der Bereich der Erfüllungserfüllung liegt zwischen ‚überhaupt nicht erfüllt‘ (-1) und ‚völlig übererfüllt‘ (+1), das mittlere Konfirmationsniveau steht für ‚gerade erfüllt‘ (0). Die Variablen der Zufriedenheitsbeiträge (ZB) stehen für die BaF, LF und BeF. Der ZB der BeF liegt im Bereich 0 bis 1. Im Bereich nicht erfüllter Erwartungen ist die Funktion nicht definiert. Der ZB der BaF liegt im Bereich -1 bis 0. Der ZB bei Nicht-Erfüllung eines LF muss mindestens so groß sein wie die der BaF, bei Übererfüllung jedoch nie so groß wie die der BeF. An den Endpunkten der Abszisse dürfen die LF nicht zu extremen ZB führen. Die Funktion liegt somit zwischen den Funktionen der BaF und BeF. Wird eine Erwartung nicht erfüllt, so wirkt sich dies, verglichen mit einer Übererfüllung, laut *prospect theory* stärker aus. Die Funktion, welche zwischen den Funktionen BeF und BaF liegt, also im Bereich der Maximalwerte +1 und -1, darf somit nur abschnittsweise linear sein. Der Wertebereich der ZB für LF, wie in Diagramm 18 dargestellt, liegt zwischen \underline{a} und \bar{a} , wobei \underline{a} größer sein muss als \bar{a} , \underline{a} zwischen -1 und 0, und \bar{a} zwischen 0 und +1 liegen muss.

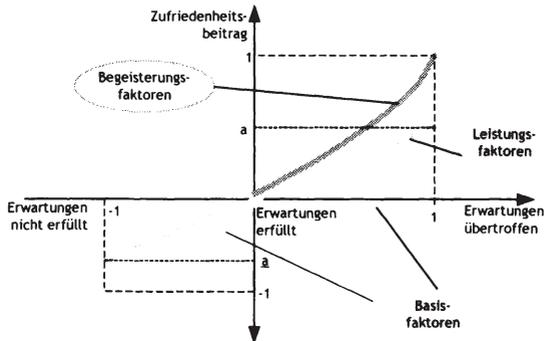


Diagramm 18: Modifiziertes Kano-Modell (Buhl et al., 2007)

Die ZB eines Kundensegmentes und einer Kategorie werden miteinander verknüpft. Dann werden die ZB der drei Kategorien zu einer Gesamtzufriedenheit für jedes Kundensegment verdichtet, wobei bei der Aggregation die Bedeutungen der Leistungen in Form der ZB segmentspezifisch gewichtet

werden. Die Art und Weise, wie einzelne Leistungen zu einem ZB pro Kategorie aggregiert werden sollen, wurde bis dato in der Literatur vernachlässigt. Das originale Kano-Modell verwendet für die Teilzufriedenheiten einen additiven Ansatz um zur Gesamtzufriedenheit zu gelangen, unterstellt jedoch, dass Unzufriedenheit in Teilbereichen, durch Zufriedenheit in anderen Bereichen ausgeglichen werden kann. Extreme Unzufriedenheit kann jedoch nur schwer ausgeglichen werden, was eine kumulative Aggregation mit Durchschnittsbildung nicht berücksichtigt. So müssen die ZB der BaF und BeF akkumulativ sein. Taucht ein zusätzlicher positiver (negativer) ZB in den Bef (BaF) auf, so müssen sich auch die aggregierten Werte erhöhen (verringern). Ist der zusätzliche ZB nämlich geringer, verglichen mit dem bereits vorhandenen ZB, so führt die Mittelwertbildung zu einem Sinken der aggregierten ZB. Bei den LF ist ein Ausgleich zwischen negativen und positiven Werten möglich, da keine extremen ZB möglich sind, welche einen nachhaltigen Einfluss auf die aggregierten ZB haben. Weiters werden LF erwartet und ein Ausgleich scheint realistischer. Somit wird für die ZB-Berechnung der LF der Mittelwert verwendet. Nach Aggregation der ZB für jede Kategorie wird die Gesamtzufriedenheit berechnet. Die Veränderung des ZB in einer Kategorie hat dabei, abhängig von den ZB in den anderen Kategorien, unterschiedliche Auswirkungen auf die Gesamtzufriedenheit. Weisen zwei Kategorien große ZB auf, so führt die Erhöhung in der dritten Kategorie zu einer stärkeren Erhöhung der Gesamtzufriedenheit, als dies der Fall wäre, wenn in den beiden Kategorien Unzufriedenheit vorzufinden wäre. Liegt der Wert der BaF über -1 so können sich die drei Kategorien ausgleichen. Falls nicht stellt der BaF ein K.O.-Kriterium für die Aggregation dar. Allgemein gehen die ZB der drei Kategorien mit unterschiedlichen Gewichtungen in das Modell ein, damit BeF den stärksten Einfluss haben und LF stärker wirken als BaF.

3.4. Konstrukte und Dimensionen höherer Ordnung

3.4.1. Wichtigkeitsdimensionen

So wie in den Zufriedenheitskonstrukten später nicht von nur einem Konstrukt ausgegangen werden kann, so gilt selbiges auch für die Wichtigkeit, worauf hier am Rande eingegangen wird. Van Ittersum et al. (2007) untersuchen die Validität unterschiedlicher Arten der Wichtigkeitsmessung. *Convergent validity* untersucht, ob unterschiedliche Messungen das gleiche Konstrukt betreffen, also beispielsweise positiv korreliert sind. *Nomological validity* untersucht, ob Messungen mit anderen Konstrukten zusammenhängen, beispielsweise die Vorhersagegenauigkeit. Zehn Messmethoden zur Attributswichtigkeit werden in Diagramm 19 aufgelistet und klassifiziert.

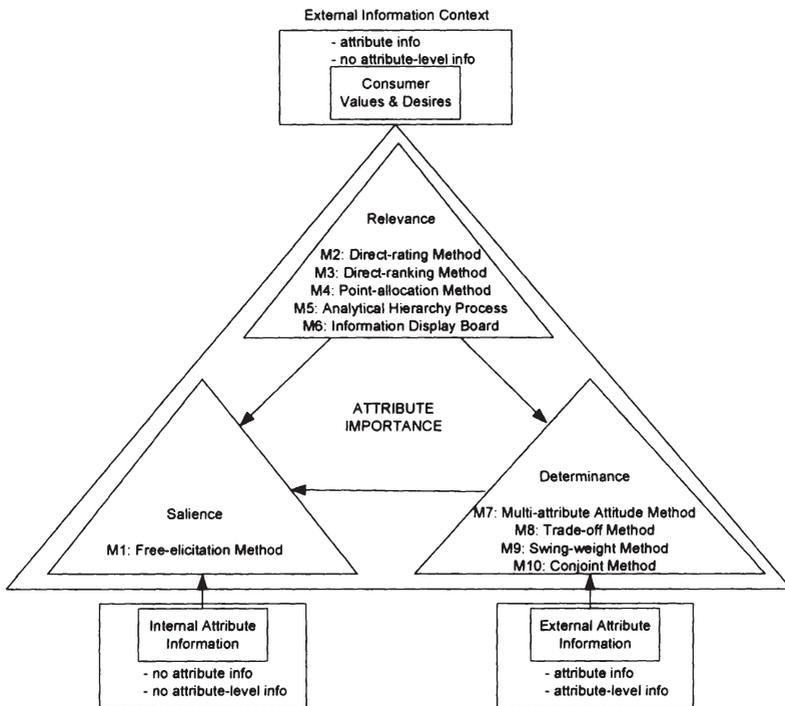


Diagramm 19: Dimensionen der Wichtigkeit (Van Ittersum et al., 2007)

M1 verwendet offene Fragen zur Angabe der wichtigen Attribute. M2 lässt die Individuen die Attribute auf einer Wichtigkeitsskala von 1 ‚unwichtig‘ bis 7 ‚wichtig‘ bewerten. M3 verlangt eine Rangordnung der Attribute. In M4 werden 100 Punkte auf die Attribute aufgeteilt. Bei M5 werden die Attribute paarweise verglichen. M6 bestimmt die Wichtigkeit anhand des Ausmaßes und der Reihenfolge der Suche danach. Bei M7 wird die Wichtigkeit durch die Regression der Gesamteinstellung auf die Attributlevelbewertung gewonnen. Bei M8 muss ein Attribut eines Produktes verändert werden, bis es gleich attraktiv wie ein Produkt mit allen vorhandenen Attributen ist. Bei M9 muss entschieden werden, welches Attribut, ausgehend vom niedrigsten Level zuerst upgraded wird und danach werden Punkte verteilt. M10 vergibt Wichtigkeiten anhand der unterschiedlichen Evaluierungen.

Ein großer Schwachpunkt in der Validität ist die Mehrdimensionalität der Attributswichtigkeit. Sie wird als unidimensionales Konzept angenommen. Hier wird jedoch ein mehrdimensionales Konzept, wobei unterschiedliche Methoden

unterschiedliche Wichtigkeitsdimensionen messen, unterstellt. Erstens werden die individuellen Wichtigkeiten der Attribute zweier Methoden korreliert und ein Wert von 0,35 wird als Grenzwert für die *convergent validity* angenommen. Zweitens wird das Verhältnis der zwei wichtigsten Attribute zweier unterschiedlicher Methoden zur *convergent validity* Bestimmung verwendet. Ist das Verhältnis nicht signifikant unterschiedlich so besteht Konvergenz. Um die *nomological validity* zu prüfen werden die Wichtigkeitswerte zweier Methoden verwendet, um die Auswahlprognose, Empfindungen, Präferenzen, oder Absichten zu prognostizieren und schließlich die Ergebnisse zwischen zwei Studien zu vergleichen. Auch die Korrelationen zwischen Attributswichtigkeit und einer der abhängigen Variablen wird verwendet. *Discriminant validity* zwischen Methoden, welche unterschiedliche Dimensionen messen, kann bestätigt werden. Der Validitätsmangel der Wichtigkeitsmessungen kann durch den multidimensionalen Ansatz erklärt werden. Es besteht *convergent validity* und *nomological validity* zwischen Methoden, welche die gleichen Wichtigkeitsdimensionen messen. Weiters ist die Korrelation zwischen Methoden gleicher Dimensionen, *convergent validity*, signifikant größer, verglichen mit den Korrelationen unterschiedlicher Dimensionen, *discriminant validity*. Für die Metaanalyse wird ein *parametric adjustability*-Modell verwendet. Dafür wird der Effekt von Dummyvariablen auf die Größe der Korrelationskoeffizienten gemessen. Die Dummyvariable ist 1, wenn die Korrelation zweier Modelle unterschiedliche Dimensionen misst, und 0, wenn sie gleiche Dimensionen misst. Der Dummyeffekt ist signifikant, sprich die Korrelationsgröße ist abhängig davon, ob die beiden Methoden gleiche oder unterschiedliche Dimensionen gemessen haben. So werden *convergent validity* zwischen, und *nomological validity* von Methoden gleicher Dimension, und *discriminant validity* zwischen Methoden unterschiedlicher Wichtigkeitsdimension verwendet.

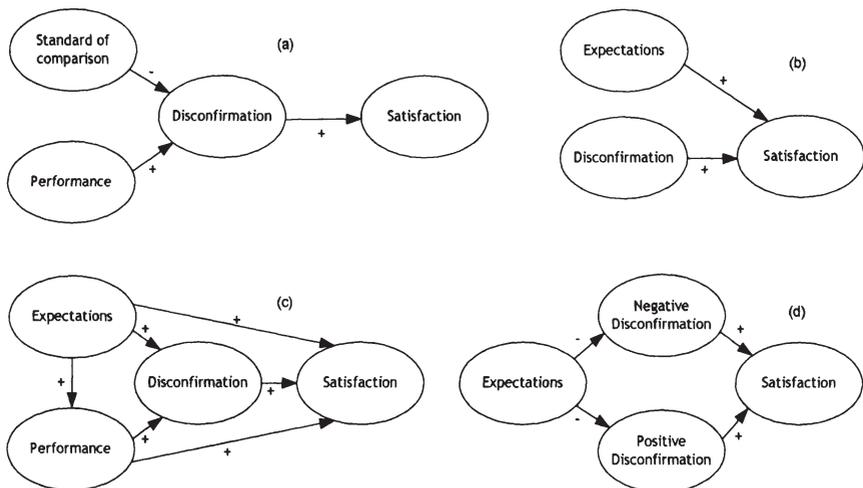
Die drei Dimensionen *salience*, *relevance* und *determinance* werden wie folgt beschrieben: 'The salience of an attribute represents the importance of the attribute in memory. The relevance of an attribute represents the importance of the attribute to the individual based on personal values and desires. Finally, the determinance of an attribute represents the importance of the attribute in judgment and choice.' Weiters wird erwähnt, dass sich abstrakte Attribute leichter evaluieren lassen, was zu einem Abstraktheitsunterschied zwischen separaten und gemeinsamen Evaluierungsaufgaben führt und die gemeinsame Evaluierung für konkrete Attribute bevorzugt. So gibt es Methoden, welche eher *relevance* und jene, die eher *determinance* messen. Zusammenfassend kann der Validitätsmangel von 67.0% auf 31.9% reduziert werden, wenn die drei

Dimensionen und die Dimensionsbeeinflusser der Wichtigkeitsattribute berücksichtigt werden. Abschließend muss festgestellt werden, dass die Wichtigkeitsdimensionalität der Klassifizierungsmethoden der Zufriedenheitsfaktoren berücksichtigt werden muss.

3.4.2. Zufriedenheitsherleitung

Bevor Zufriedenheitsdimensionalitäten angesprochen werden, muss zuerst die Herleitung der Zufriedenheitskonstrukte veranschaulicht werden. Campo und Yagüe (2009) beschreiben den Einfluss der vorangehenden empfundenen Qualität und des Preises auf die Zufriedenheit der Konsumenten beim Kauf. Es zeigt sich, dass die empfundene Qualität einen positiven signifikanten Einfluss auf die Zufriedenheit hat. Dieser Effekt ist größer als der des empfundenen Preises. Der gesamte Preiseffekt wird auf zwei Teile aufgesplittet. Ein nicht-linearer negativer Effekt, welcher das Opfer des empfundenen Preises betrifft, also dem Modell der *decreasing returns* folgt, und ein positiver Effekt, welcher den Einfluss des Preises als Signal der Qualität repräsentiert und durch die Spanne des minimalen und maximalen Preises, den ein Konsument als akzeptabel ansieht, repräsentiert wird. Den Einfluss der Vorgänger vor der Zufriedenheit zu analysieren ist wichtig, da dadurch der Einfluss von Verbesserungen der Servicequalität auf die Zufriedenheit evaluiert werden kann. Die Preislevels zur Steigerung der Zufriedenheit können bestimmt und der positive Einfluss der empfundenen Qualität kann gelenkt werden. So ergeben sich unterschiedliche Konzeptionalisierungsmöglichkeiten der Zufriedenheit. Zufriedenheit kann als gesamte Evaluierung nach dem Kauf gesehen werden. Der psychologische Status, welcher die umschließende Emotion der Erwartungen ausschließt, kann mit den vorangehenden Gefühlen der Konsumation gekoppelt werden. Es kann verglichen werden was jemand bekommen hat, mit dem was jemand erwartet hat. Auch der Nutzen und die Kosten können verglichen werden. Zufriedenheit kann auf eine spezielle Transaktion und auf die globale Zufriedenheit mit dem Service, oder auch als kumulierte Zufriedenheit, welche von allen vorangehenden Käufen abhängt, beschrieben werden. Der Nutzen-Kosten-Ansatz fokussiert die kognitive Komponente. Aber auch die emotionale Komponente spielt eine Rolle in der Zufriedenheit, weshalb beide Komponenten miteinbezogen werden müssen. Diese können sich wiederum auf einen speziellen oder auf die kumulierten Käufe beziehen. Die emotionale Komponente kann als ein Resultat eines speziellen Kaufes oder einer Abfolge von Käufen gesehen werden. Zufriedenheit entsteht zum Beispiel aus einem Vergleich zwischen den Erwartungen vor dem Kauf und der Wahrnehmung im Zuge der Konsumtion. Dies wird als Disconfirmation beschrieben, wobei diese positiv ist, wenn die Wahrnehmung besser ist als die

Erwartung, null wenn die Wahrnehmung den Erwartungen entspricht und negativ wenn die Wahrnehmung schlechter als die Erwartung ist. Diagramm 20 zeigt die unterschiedlichen Konstrukte der Zufriedenheit.



Sources: (a) Cadotte et al. (1987); (b) Oliver (1981); (c) Bolton and Drew (1991); Yi (1991); Yi and La (2003); (d) Adaptation of Anderson and Sullivan (1993)

Diagramm 20: Zufriedenheitskonstrukte (Campo und Yagüe, 2009)

Sie kann direkt positiv über Disconfirmation beeinflusst werden (a). Disconfirmation und Erwartung können die Zufriedenheit beeinflussen, wobei der Einfluss der Erwartungen geringer ist als der der Disconfirmation (b). Zusätzlich können die Wahrnehmungen die Zufriedenheit direkt beeinflussen, wobei die Erwartungen und die Wahrnehmungen einen indirekten Effekt über Disconfirmation ausüben (c). Schließlich kann zwischen positivem und negativem Disconfirmationseinfluss auf die Zufriedenheit unterschieden werden, wobei letzterer den größeren Effekt hat (d).

Aufgrund der Unstimmigkeiten wird die Zufriedenheit hier aus einem unterschiedlichen theoretischen Fokus heraus erklärt. Die *satisfaction theory* stammt aus der Nutzentheorie, wobei die Bestimmungsfaktoren der Zufriedenheit, also der wahrgenommene Gewinn, meist als wahrgenommene Qualität gemessen werden. Das Opfer, welches erbracht werden muss, um das Produkt zu bekommen, berücksichtigt monetäre und nicht-monetäre Kosten. Relevant ist dabei die Beziehung zwischen wahrgenommenem Preis und Zufriedenheit. Die Spanne des akzeptablen Preises tritt als Referenzpreis hervor, welchen der Konsument als Indikator für die Spanne der wahrgenommenen

Qualität sieht, also vom minimalen bis zum maximalen Qualitätslimit. So wird einerseits der negative Preiseffekt als Indikator für das Opfer, welches erbracht werden muss, und andererseits der positive Effekt der Spanne des akzeptablen Preises als Indikator für Qualität gemessen. Es wird davon ausgegangen, dass eine negative nicht-lineare Beziehung zwischen Preis und Zufriedenheit besteht. Getestet wird, ob diese negative nicht-lineare Beziehung eine sinkende Wachstumsrate, oder eine umgekehrte U-Form aufweist. Letztere würde bedeuten, dass niedrige und hohe Preise zu geringerer Zufriedenheit führen und mittlere Preise zu höherer Zufriedenheit. Die Hypothese mit sinkenden Erträgen kann bestätigt werden.

3.4.3. Delight

Torres und Kline (2006) grenzen Zufriedenheit und *delight* voneinander ab. Unter *delight* wird das Übertreffen von Zufriedenheit verstanden, was zu größerer Emotion und somit zu einem unterschiedlichen psychologischen Status der Zufriedenheit führt. Zufriedenheit bedeutet die Erwartungen zu erfüllen, *delight* hingegen diese zu übertreffen. Die Konsumenten weisen einen Zufriedenheitsbereich der *zone of tolerance* genannt wird auf. Wird dieser an der oberen Grenze übertroffen, so führt dies zu *delight*. ‚Sehr zufrieden‘ bedeutet aber nicht *delight*, da letzteres als stärkerer Status der Einstellung verstanden wird. *Delight* wird als Kontinuum erhöhter Erwartung angesehen und Zufriedenheit als Vorgänger. Ob Zufriedenheit und *delight* unabhängig sind oder nicht, wurde noch nicht vollends geklärt. Manche sehen *delight* als Ergänzung von Zufriedenheit an. Im Hinblick darauf, dass *delight* eine Funktion der Erwartung ist, so muss auch das Zufriedenstellen des Kunden ein Vorbote von *delight* sein. Aus der Sichtweise von Bedürfnissen wird Zufriedenheit als Erfüllung von bestimmten Bedürfnissen angesehen (z.B. Sicherheit), *delight* hingegen mit der Erfüllung höherer Bedürfnisse (z.B. Selbstachtung). So werden *delight* und Zufriedenheit als unterschiedliche Konzepte mit einer korrelierten Beziehung verstanden. Sie beschreiben unterschiedlich emotionale Stadien in Bezug auf Kundenzufriedenheit und Servicequalität.

Diagramm 21 zeigt den Übergang von Zufriedenheit zu *delight*. Damit Konsumenten zufrieden sind, müssen ihre Erwartungen erfüllt werden. Werden die Eigenschaften auf der linken Seite erfüllt, steigt die Einstellung von *satisfied* auf *very satisfied*. Einige der Elemente können *delight* hervorrufen, was einer stärkeren Einstellung, verglichen mit *very satisfied*, entspricht. Die übertroffenen Erwartungen mit den entsprechenden Eigenschaften auf der rechten Seite bewirken *delight*. Unterscheidet sich die Erfahrung von den notwendigen Bedürfnissen stark, so kann die Erfüllung eines Bedürfnisses höherer Ordnung

emotionale *delight*-Gefühle auslösen. Zufriedenheit wird zusammengefasst als mehrdeutiger emotionaler Status beschrieben, wobei *delight* einen höheren Grad an Engagement und emotionaler Erregung auf der positiven Seite mit sich bringt.

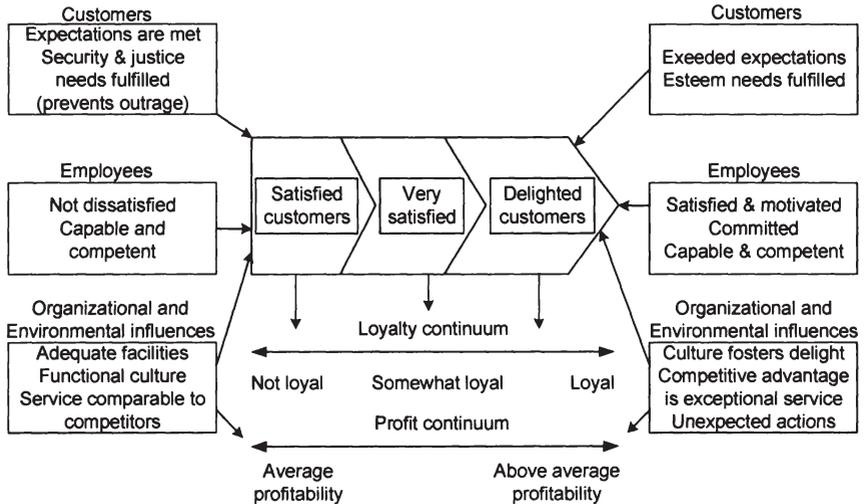


Diagramm 21: From Satisfaction to delight (Toores und Kline, 2006)

Was unter *delight* genau zu verstehen ist, untersucht Finn (2005). Er stellt die Frage ob *delight* mehr als ein bloßer nicht-linearer Effekt der Zufriedenheit auf die Absichten ist. Wenn die Zufriedenheit die Schwelle der *zone of tolerance* überschreitet; hat die Serviceerfahrung einen Einfluss auf *delight*. Der Zufriedenheitseffekt, welcher auf die Loyalitätsmerkmale wirkt, nimmt ebenfalls in höheren Zufriedenheitslagen zu. Auch gegenteiliger Effekt kann gezeigt werden, nämlich abnehmende Effekte in höheren Zufriedenheitslagen. Zufriedenheit wird als zusammenfassende Evaluierung angesehen, aber auch dazu herrscht Unklarheit, da diese primär als kognitiv angesehen wird, jedoch auch schon emotional konzeptualisiert wird. Eintreffende Erwartungen führen zu Zufriedenheit, jedoch erst die emotionale Antwort auf eine Überraschung und einen positiven Level von Performance zu *delight* oder *outrage* und schließlich zu Loyalität. In Diagramm 22 stellen *delight* und Zufriedenheit unterschiedliche Konzepte dar, welche diese Absichten beeinflussen. Die vorangehende Kausalkette für *delight* liegt in unerwartet hohen Levels an Performance oder Zufriedenheit, also eine überraschende Konsumerfahrung, welche *arousal* auslöst und wiederum zu *delight* führt.

Mit diesem Modell soll festgestellt werden, ob der Effekt von *delight* auf *intention* mehr als nur ein nicht-linearer Effekt der Zufriedenheit auf *intention* ist. Pfadmodelle um den linearen Einfluss von Zufriedenheit und *delight* auf *intention* zu messen, zeigen Inkonsistenzen des Effektes. Gründe dafür können in der Nicht-Linearität des Effektes Zufriedenheit auf *intention* liegen, welcher selbst inkonsistent ist. So wird in vorgestelltem Modell ein quadratischer Effekt der Zufriedenheit eingeführt. Weiters wird ein Modell getestet, welches Interaktion zwischen Zufriedenheit und *delight* erlaubt. Es wird untersucht, ob quadratische oder Interaktionseffekte der Zufriedenheit Auswirkungen auf den Effekt von *delight* auf *intention* haben.

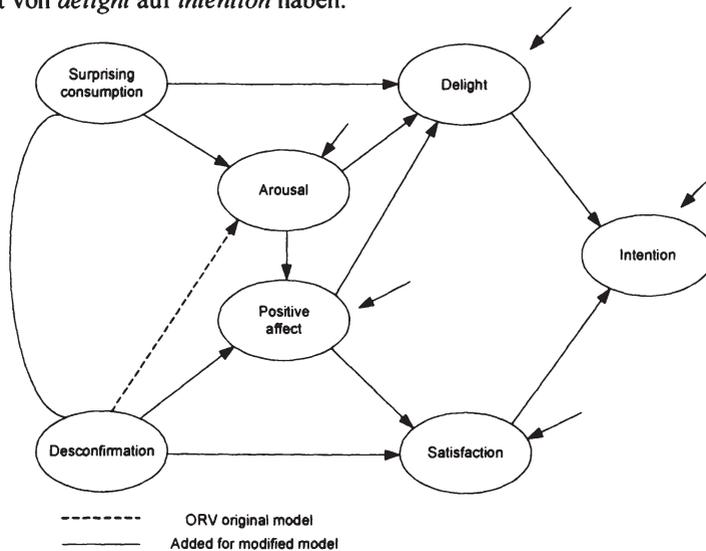


Diagramm 22: Delight and Satisfaction (Finn, 2005)

Weder der quadratische noch der Interaktionseffekt erreicht einen ausreichenden Signifikanzlevel. Werden der schwach positiv quadratische Effekt, oder beide schwachen Effekte, also der quadratische und der negative Interaktionseffekt berücksichtigt, so sinkt der lineare Effekt der Zufriedenheit auf *intention* und der lineare *delight* Effekt wird signifikant. Dies bestätigt die konzeptionelle Unterscheidung zwischen *delight* und Zufriedenheit von Services mit höherem Grad an *involvement*, nun auch für das alltägliche Kundenservice. So wird die Behauptung nicht unterstützt, dass *delight* nur einen signifikanten Einfluss auf *intention* in Servicebereichen mit höherem *involvement* hat. Alle Pfade in Diagramm 22 können bestätigt werden. *Delight* ist nicht nur ein nicht-linearer Effekt der Zufriedenheit auf *intention*, sondern unterstützt auch die emotionale

vs. kognitive konzeptionelle Unterscheidung zwischen *delight* und Zufriedenheit. *Delight* wird als getrennte Erwidern eines Service gesehen. Die Streuung von *delight* kann erklären, warum Konsumenten mit gleichem Zufriedenheitslevel unterschiedliche Verhaltensintentionen aufweisen. Die Behauptung, dass *delight* bloß der nicht-lineare Effekt der Zufriedenheit auf *intention* ist, wird nicht unterstützt. Die Diskriminanzvalidität des *delight* und Zufriedenheitskonstruktes separiert die emotionalen und kognitiven Antwortverhalten und beeinflusst die Verhaltensabsichten.

Arnold et al. (2005) beschreiben warum *delight* in der Zufriedenheitsforschung an Wichtigkeit gewinnt und in zahlreichen verhaltensspezifischen Untersuchungen deren Auswirkung gemessen wird. *Delight* wird als extrem wichtig in Bezug auf die langfristige Loyalität gesehen, da die unerwartet erfreulichen Erfahrungen höhere Emotionsebenen als Zufriedenheit wecken. Hergeleitet wird die Existenz ausgehend vom *Expectation-Disconfirmation-Modell*. Übertrifft die Performance die Erwartungen, so entsteht positive Diskonfirmation. Diskonfirmation wird in drei Kategorien aufgrund der Unerwartetheit unterschieden. Konfirmation mit leichter Abweichung wird als normal betrachtet. Diskonfirmation wirkt plausibel, solange sie nicht häufig auftritt. Diskonfirmation, die unerwartet, überraschend und unwahrscheinlich aufgrund vergangener Erfahrung ist, löst *surprise disconfirmation* aus. Dies bildet die kognitive Basis für *delight*. *Delight* stellt jedoch nicht bloß eine kognitive Beurteilung dar. Auch die affektive Komponente, die *second-level* Emotion als Kombination von Emotionen auf niedrigerem Level, spielt dabei eine Rolle. So ist *delight* eine Funktion aus positiver Diskonfirmation, Erregung und positivem Affekt. Das Erreichen dieser Ebene führt zu höheren zukünftigen Erwartungen. Dies macht das langfristige Erreichen dieser Stufe für alle Konsumenten, aufgrund der notwendig höheren Performance, kaum möglich.

3.4.4. Kognitive und affektive Beeinflussung der Zufriedenheit

Ladhari (2007) beschreibt die Zufriedenheitsherleitung im Kinobesucherbereich. Das Kinoerlebnis als intangibles Objekt, wird aus *pleasure-* und *arousal-*Gründen besucht und nicht zwecks wirtschaftlicher Gewinnmaximierung und ist somit als empirisches Beispiel bestens tauglich. Er sieht Diskonfirmation affektiver Erwartungen als signifikanten Vorgänger der Zufriedenheitsbewertungen und baut sein Modell über das *Expectation-Disconfirmation-Modell* auf, welches einen kognitiven Prozess hinsichtlich vorangehender Erwartungen und empfundener Performance beschreibt. Zusätzlich bringt er eine affektive Komponente ein. Er geht von einem, vom kognitiven Prozess abhängigen oder unabhängigen affektiven Prozess, also einer zweidimensionalen

Struktur der Diskonfirmation der Erwartungen aus. So unterstellt er zwei direkt positive Beziehungen der Diskonfirmation von kognitiven Erwartungen und Diskonfirmation von affektiven Erwartungen mit der Zufriedenheit auf. Zusätzlich beeinflussen *pleasure* und *arousal* die Zufriedenheit positiv, welche laut Russell's *model of affect* die emotionalen Kaufreaktionen einer Person auffangen. *Pleasure* steht für den Grad an Wohlgefühl und *arousal* für den Grad der Stimulierung. Weiters spielen die Diskonfirmation von kognitiven Erwartungen und Diskonfirmation von affektiven Erwartungen in der Emotionsempfindung eine Rolle. Wird die Performance besser als erwartet empfunden, so entstehen positive Emotionen, ist sie schlechter führt dies zu negativen Emotionen. So wird der Einfluss beider Diskonfirmationsebenen auf beide Emotionsebenen unterstellt. Abgesehen vom Einfluss der Diskonfirmation affektiver Erwartungen auf *pleasure* bestätigt ein verbesserter Modellfit die Annahmen. In Diagramm 23 werden zusätzlich die positiven und die negativen Emotionen aufgeteilt, welche positive Effekte auf die Zufriedenheit zeigen.

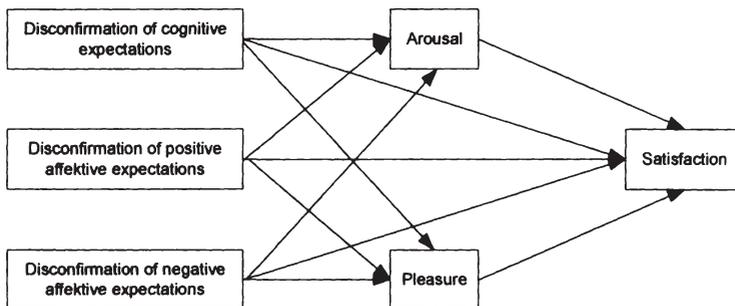


Diagramm 23: Revised model (Ladhari, 2007)

Liegt die kognitive Performance unter den Erwartungen, sind die Kinogehrer trotzdem zufrieden, solange die negative Diskonfirmation durch positive Diskonfirmation der affektiven Erwartungen kompensiert wird. Auch Diskonfirmation negativ affektiver Erwartungen hat einen Einfluss auf die Zufriedenheit. Negative Emotionen steigern also den Zufriedenheitslevel, wenn diese der Häufigkeit der erwarteten negativen Emotionen entsprechen oder übersteigen. Ist die Häufigkeit negativer Emotionen niedriger als erwartet, wird Freude empfunden, jedoch weniger intensiv mit niedrigerer Zufriedenheitsempfindung. Negative Emotionen beeinflussen den Zufriedenheitslevel positiv. Diese Ergebnisse unterstreichen die Wichtigkeit positiver und negativer Emotionen in der Zufriedenheitsmessung.

3.4.5. Berücksichtigung der SERVQUAL Faktoren

Kueh (2006) liefert ein Beispiel zu *satisfiers* und *dissatisfiers*. Kritisiert wird die implizite Annahme der Konstrukte gemäß SERVQUAL, dass die gleichen Faktoren, welche Zufriedenheit verursachen, auch Unzufriedenheit verursachen. Zufriedenheit und Unzufriedenheit als zwei gegenteilige Enden des gleichen Konstrukts zu sehen ist jedoch fraglich. Auch Herzberg trennt zwei separate Sets an Faktoren auf und sieht diese nicht als Gegenteile auf einem bipolaren Kontinuum. Die instrumentelle Performance wird als Hygienefaktor bezeichnet, welche Unzufriedenheit auslösen kann, aber nicht zu Zufriedenheit führt. Zufriedenheit wird durch die expressive Performance ausgelöst, als auch durch die instrumentelle. Auch in anderen Studien, die Faktoren in *dissatisfiers*, *satisfiers*, *criticals* und *neutrals* einteilen, wird *dissatisfiers* die Rolle der Hygienefaktoren zugesprochen. Es gibt jedoch auch *satisfier*, welche mit intangiblen interpersonellen Aspekten verbunden sind und *dissatisfier*, welche mit operationellen Aspekten verbunden sind. So wird untersucht, ob unterschiedliche Sets an Faktoren für Zufriedenheit bzw. Unzufriedenheit verantwortlich sind.

Satisfiers und *dissatisfiers* werden mit der *critical incident technique* (CIT) identifiziert. Im Anschluss wird untersucht, ob diese aus den gleichen Faktoren bestehen, oder nicht. Die CIT ermöglicht den Probanden, offen Geschichten extrem zufriedener und unzufriedener Serviceerfahrungen zu erzählen. Zwei Beurteiler teilen die Geschichten in zufriedene und unzufriedene Erzählungen ein. Die Zuteilung nach Korrektur bei Ungereimtheiten ergibt folgende allgemeine Einteilung: *criticals*, *satisfiers*, *dissatisfiers* und *neutrals*. Bestimmte Servicedimensionen werden Es entweder signifikant positiven oder negativen Ereignissen zugeteilt. Zwischen *satisfiers* und *dissatisfiers* ergibt sich anstatt einer negativen Korrelation eine extrem schwache. So sind Faktoren, welche Zufriedenheit auslösen, nicht unbedingt die gleichen die Unzufriedenheit bewirken, und vice versa. Die Ergebnisse unterscheiden sich zwischen den unterschiedlichen Servicebereichen. Auch die Wichtigkeit der Servicequalitätsdimensionen in Bezug auf die Zufriedenheit ist abhängig vom Servicebereich.

3.4.6. Eine polynomiale Lösung

Mazanec (2006) untersucht die Auswirkungen latenter Konstrukte mehrerer reflektiver Indikatoren auf die Gesamtzufriedenheit. Anhand einer Clusteranalyse mit TRN (Topology Representing Network) kann die Unidimensionalität des Zufriedenheitskonstruktes gezeigt werden. Die Gruppennummern in Diagramm 24 zeigen in geordneter Reihenfolge (4, 3, 2, 1, und 5) die sinkende

Zufriedenheit des Gesamtzufriedenheitskonstruktes zweiter Ordnung. Die fünf Zufriedenheitsgruppen werden mittels Sammon mapping aus einem acht-dimensionalen Raum auf einen zwei-dimensionalen Raum projiziert.

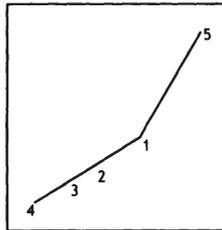


Diagramm 24: Ordered satisfaction clusters (Mazanec, 2006)

Es wird davon ausgegangen, dass die acht bereichsspezifischen Faktoren die drei Kano-Faktoren (*ordinary*) *performance*, *hygiene* und *excitement* Faktoren enthalten. Durch die nicht-lineare Abweichung von der linearen Beziehung zwischen Bereichsfaktoren und der Gesamtzufriedenheit, soll die Existenz von *excitement (delight)* und *hygiene factor (dissatisfier)* gezeigt werden. Die *ordinary performance factors* sollten keine signifikanten nicht-linearen Effekte zusätzlich zu ihrem linearen Effekt zeigen. Um die Asymmetrie für niedrige und hohe Werte zu bestimmen, wird ein Polynom dritter Ordnung verwendet und für jeden Faktor simultan der lineare, der parabolische und der kubische Term bestimmt. Es zeigt sich, dass die Einführung eines quadratischen und eines kubischen Effektes die Stärke und Größe der linearen Parameterwerte nicht zerstört, was die mögliche Trennung der linearen und nicht-linearen Komponenten bestätigt. Zusätzlich steigt der Modellfit in jenen Modellen, in welchen der nicht-lineare Term signifikant ist. Die quadratischen und kubischen signifikanten Terme sind in Diagramm 25 einsehbar.

Die Ergebnisse stimmen nicht mit den zuvor aus dem Kano-Modell nach der Herzberg-Theorie postulierten Annahmen überein. ‚Services‘ stellt einen Hygienefaktor dar. ‚Restaurants‘ weckt den Anschein, als ob das Attribut ein Begeisterungsfaktor wäre. In der rechten Ecke ist jedoch erkennbar, dass dieser Faktor Unzufriedenheit auslöst, wenn er nicht ausreichend befriedigt wird. So wird die Zufriedenheitstheorie um die Annahme erweitert, dass es Faktoren gibt, welche starke Unzufriedenheit bei schlechter Performance auslösen und gleichzeitig Enthusiasmus auf höheren Performancelevels auslösen.

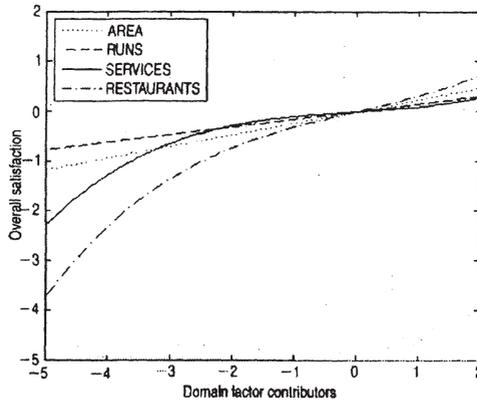


Diagramm 25: Contributions to Overall Satisfaction (Mazanec, 2006)

Mazanec (2007) liefert zusätzlich Hinweise, um bei der Suche nach Satisfaktoren und Hygienefaktoren Missbrauch in Strukturgleichungsmodellen zu vermeiden. Er hinterfragt die Annahme, ob sozialwissenschaftliche Querschnittsuntersuchungen nur assoziative, jedoch keine kausalen Zusammenhänge aufzeigen können, was im SEM-Bereich vermieden wird. Aus der *inferred causation theory* leitet er einen erweiterten Interpretationsspielraum aus Querschnittsdaten ab. Die *trial&error*-Modellanpassung, über Modifikationsindizes zwecks Steigerung des Modell-Fits im SEM Bereich, wird im Nachhinein oft als Modelltest verkauft. So wird die Verwendung einer zufälligen Teilstichprobe empfohlen, sofern kein unbearbeiteter Datensatz zur Verfügung steht. Trotz unterschiedlicher Relationssysteme kann oft die gleiche beobachtete Kovarianzmatrix reproduziert werden, was SEM als konfirmatorischen Ansatz in Frage stellt. Auch die Verwendung formativer Indikatoren, welche eine latente Variable erzeugen vs. reflektive Indikatoren, welche Manifestationen der Latenten sind, wird häufig nicht theoretisch-konzeptionell argumentiert und führt zur Verwendung der einfacher handhabbaren reflektiven Indikatoren. Auch die (Intervall)Skalen- und Normalverteilungsannahmen wurden durch zahlreiche Verwendung von Ratingskalen oft verletzt. In neuerer Zeit wird diesem Problem mit Hilfe von robusten Schätzern für ordinale, also geordnet kategoriale, oder numerische Variablen Abhilfe geschaffen. Es wird eine normalverteilte latente Hintergrundvariable unterstellt und für diese Schwellenwerte, *thresholds*, geschätzt.

Die genannten Argumente werden anhand des Auftretens nicht-linearer Beziehungen zwischen Zufriedenheitsfaktoren und Gesamtzufriedenheit, als auch Loyalität und Wiederkaufabsicht dargestellt. Um nicht-lineare Wirkungen

aufzudecken werden gleichzeitig lineare und quadratische Annahmen spezifiziert. Bei Begeisterungsfaktoren muss der Koeffizient des quadratischen Terms signifikant positiv sein, bei Dissatisfaktoren signifikant negativ und bei Leistungsfaktoren insignifikant. Wie in Diagramm 26 einsehbar, können Dissatisfaktoren, als auch Leistungsfaktoren aufgefunden werden.

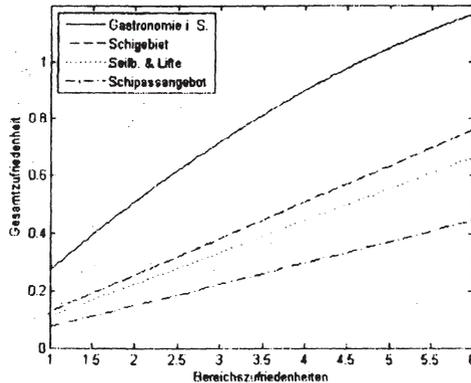


Diagramm 26: Wirkungsbeiträge von vier ausgewählten Zufriedenheitsfaktoren (Mazanec, 2007)

Um die Robustheit zu prüfen wird die metrische mit der nicht-metrischen Lösung verglichen. Werden die Indikatoren als kategorial betrachtet, so wird diesen eine normalverteilte latente Variable zugrunde gelegt und geeignete *thresholds* für die Ratings geschätzt. Die Koeffizienten der nicht-metrischen Lösung sind der metrischen ähnlich. Eine weitere Prüfung wird mit Hilfe des *split-half* Verfahrens durchgeführt. Die gesamte Stichprobe wird zufällig in zwei Hälften geteilt und die metrische Variante gerechnet. Aufgrund der Stichprobengröße wird eine Übereinstimmung der Ergebnisse erwartet, welche sich bis auf Ausnahmen auch erfüllt. Die Wirkungsrechnung bei den schwachen Faktoren (Pisten, Services) kann jedoch nicht reproduziert werden, da der nicht-lineare Term die niedrige lineare Wirkung nach oben korrigiert. Stünde nur eines der beiden Teilsamples zur Verfügung, würden voreilige Rückschlüsse bezüglich der Begeisterungsfaktoren zu fraglichen Interpretationen führen. Abschließend werden konzeptionelle Erweiterungen, wie beispielsweise *hybrids*, welche bei niedrigen Werten sich wie Begeisterungsfaktoren verhalten und anhand kubischer Terme darstellbar wären, angesprochen. Weiters wird angemerkt, dass die Entfernung der nicht-signifikanten Zufriedenheitsfaktoren aus dem Modell unter erneuter Schätzung zur Herauskrystallisierung sämtlicher Zufriedenheitsfaktortypen führt.

3.4.7. Faktoreneinteilung unter zusätzlichen Annahmen

Pollack (2008) stellt eine Analysemöglichkeit zur Unterteilung der Attribute in die unterschiedlichen Kategorien vor. *Satisfiers* weisen zu Beginn keine Beziehung zur Zufriedenheit auf und erst nach Überschreitung eines Qualitätslevels, dem *inflection point*, sind sie positiv verbunden. *Dissatisfiers* haben zu Beginn eine positive Beziehung zur Zufriedenheit, aber nach dem *inflection point* keine bzw. eine sehr schwache. Auch die unendliche Erhöhung der Servicequalität muss in Frage gestellt werden. Bei *dissatisfiers* gibt es nach Erreichen des akzeptablen Qualitätslevels in der Zufriedenheit keinen Zugewinn mehr. Bei *satisfiers* muss erst eine Grenze überschritten werden, bevor Zufriedenheitsgewinne entstehen können.

Es finden sich zwei theoretische Strömungen. Die *zone of tolerance* bestimmt zwei Grenzen, manchmal auch nur eine, wobei die lineare Beziehung zwischen Qualitätsattributen und Zufriedenheit zweifelhaft ist. Der zweite Strom stammt aus der Herzberg-Zwei-Faktoren-Theorie. Diese teilt die Attribute in zwei Kategorien, wobei eine bei Erfüllung Zufriedenheit auslöst, *motivators*, und die andere bei Nichterfüllung Unzufriedenheit, *hygiene factors*. Oftmals existiert diese Trennung in zwei Faktoren aufgrund von Überlappungen nicht. Auch die Nichtlinearität, ausgelöst durch sinkende Erträge aufgrund des progressiven Verlaufes, wird oftmals nicht beachtet. *Criticals* beschreiben Attribute, welche einen linearen Verlauf annehmen. Eine multidimensionale Aufteilung der Servicequalität nach *outcome quality*, *interaction quality* und *physical service environment quality* wird verwendet. *Tangibles* aus dem SERVQUAL Ansatz entsprechen dem *physical environment aspect*, *reliability* dem *outcome aspect*, und die verbleibenden, *responsiveness*, *empathy* und *assurance*, der *interaction quality*. Die *zone of tolerance* wird durch die adäquaten und die erwünschten Erwartungen bestimmt, wie Diagramm 27 zeigt.



Diagramm 27: Zone-of-tolerance (Pollack, 2008)

Empirische Beobachtungen können nur den Übergangspunkt des akzeptablen Levels auffinden. Weitere Untersuchungen trennen *attractive, one-dimensional, must-be* und *indifferent quality* Elemente voneinander ab, oder *satisfiers, dissatisfiers, criticals* und *neutrals*. *Criticals* stellen *top satisfiers* oder *top dissatisfiers* dar. Dies wird in Diagramm 28 visualisiert.



Diagramm 28: Acceptable level (Pollack, 2008)

Es wird unterstellt, dass nicht alle Servicequalitätsdimensionen einem linearen Pfad folgen, so wie die *criticals*, und einige Dimensionen in *satisfiers* und *dissatisfiers* eingeteilt werden können. Dafür wird ein Regressionsmodell in Formel 3 formuliert.

$$Sat = \beta_0 + \beta_1 SQ + \beta_2 d + \beta_3 dSQ + \varepsilon$$

Formel 3: Regressionsmodell (Pollack, 2008)

Sat steht für Zufriedenheit, *SQ* für den Servicequalitätssubdimensionswert, *d* ist eine Dummyvariable mit Wert 1, wenn oberhalb des Akzeptanzlevels der Qualität der Subdimension, und anders 0, β_x ist der standardisierte Regressionskoeffizient und ε der Errorterm. β_1 ist der *slope* unterhalb des akzeptablen Qualitätslevels und somit der durchschnittliche Servicequalitätsdimensionseffekt auf die Zufriedenheit. β_3 ist der zusätzliche Beitrag über dem akzeptablen Level. $\beta_1 + \beta_3$ ist somit die *slopes* über dem akzeptablen Qualitätslevel. Eine Servicequalitätssubdimension wird als *critical* klassifiziert, wenn der standardisierte *beta*-Koeffizient einer Dimension signifikant positiv ist und der standardisierte β -Koeffizient der Servicequalitätssubdimension/Dummy-Variable-Interaktion nicht signifikant ist, *satisfier*, wenn ersterer nicht signifikant und letzterer signifikant positiv ist, oder beide signifikant positiv sind, *dissatisfier*, wenn ersterer signifikant positiv und letzterer negativ ist. Daraus wird geschlussfolgert, dass es zwei Nichtlinearitätstypen gibt. Vor und nach dem akzeptablen Level werden die Attribute in *satisfier* oder *dissatisfier* klassifiziert. Dieser existiert in Abhängigkeit des Serviceattributes und dem Servicetyp. Viele Attribute sind *criticals*, was die Herzberg Theorie wiederum in Frage stellt.

Es wird angemerkt, dass sich die Eigenschaften *satisfier*, *dissatisfier* oder *critical* aufgrund von sich verändernden Erwartungen über die Zeit ändern können und zusätzlich die ermittelten Erwartungen nicht die gesamte Population repräsentieren.

3.4.8. Nichtlinearitäten und Interaktionen

Nicht-Linearitäten und Interaktionseffekte in Strukturgleichungsmodellen stellen einen enormen Rechenaufwand dar. Klein und Muthén (2007) beschreiben einen Algorithmus, welcher den Schätzaufwand einschränkt und gleichzeitig beide Effekte zulässt. Sie liefern dazu ein empirisches Beispiel. Erstellt wird ein Interaktionsmodell mit zwei latenten Produkttermen und sieben Indikatorvariablen. Untersucht wird der Effekt von subjektiv empfundener Fitness, objektiver Fitness und Flexibilität der Zielerreichung auf den Beschwerdelevel der mentalen oder physischen Situation. Die Interaktionshypothese unterstellt, dass der lineare Effekt der Flexibilität der Zielerreichung auf den Beschwerdelevel hoch ist, wenn niedrige Werte in der Fitnessskala vorherrschen, dies jedoch nicht durch hohe Werte auf der Fitnessskala neutralisiert wird. Personen mit hoher subjektiver Fitness wird ein geringer Effekt der Flexibilität der Zielerreichung auf den Beschwerdelevel unterstellt. Personen mit gering wahrgenommenen körperlichen Ressourcen wird ein hoher Effekt der Flexibilität der Zielerreichung auf den Beschwerdelevel unterstellt. Diese Interaktionshypothese wird durch einen Produktterm im SEM modelliert. Die *scatterplots* in Diagramm 97 zeigen die Faktorenwerte der Flexibilität in der Zielerreichung und die beobachteten Werte des Beschwerdelevels der Individuen. In der linken Graphik befinden sich jene mit niedrigen Faktorenwerten subjektiver Fitness und in der rechten Graphik jene mit hohen Werten. Zusätzlich werden die Regressionslinien eingezeichnet. Die subjektiv empfundene Fitness zeigt sowohl einen höheren linearen, also auch interaktiven Effekt, verglichen mit der objektiven Fitness.

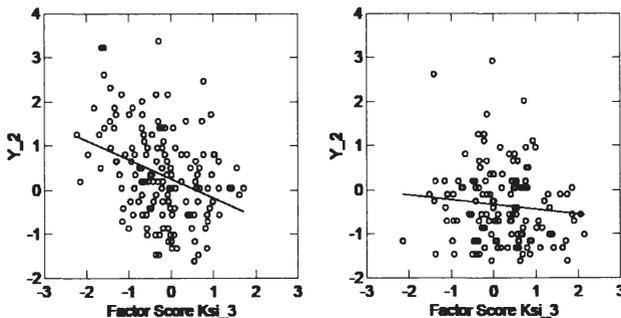


Diagramm 29: Faktorscores (Klein und Muthén, 2007)

Paulssen und Sommerfeld (2006) liefern ein empirisches Beispiel für die Einführung von Nichtlinearitäten in Strukturgleichungsmodellen, welches den hier vorliegenden theoretischen Schwerpunkten entspricht. Trotz hoher Zufriedenheit sind Kundenverluste zu verzeichnen. Der Zusammenhang zwischen Zufriedenheit und Loyalität wurde als linear und symmetrisch angenommen, wird jedoch aufgrund der fraglichen Annahme, dass Zufriedenheitssteigerungen in hohen Zufriedenheitsbereichen die gleiche Auswirkung haben wie in niedrigen Zufriedenheitsbereichen, laut Kahneman & Tversky's *prospect theory*, in Frage gestellt. Diese berücksichtigt die Auffassung von Gewinnen und Verlusten in Bezug auf einen Referenzpunkt. Unter dem Begriff *loss aversion* wird die Verlustfunktion, verglichen mit der Gewinnfunktion, als steiler angenommen. So führen Gewinne und Verluste gleicher Stärke zu unterschiedlichen Entscheidungseinflüssen. Verluste wirken sich stärker aus als Gewinne und negative Attributperformance hat mehr Gewicht in Bezug auf die Wiederkaufwahl, verglichen mit positiver.

Methodisch geben Paulssen und Sommerfeld (2006) SEM den Vorzug vor multivariaten Regressionsanalysen. Loyalität und Zufriedenheit sind reflektive latente Konstrukte, wobei letztere Modelle nicht die Messfehler korrigieren und dies zu inkonsistenten und schwachen Parameterschätzungen führt. Weiters führt die Dummymodierung zu Informationsverlust. Als Lösungsvorschlag verwenden sie Klein's SEM mit quadratischen latenten unabhängigen Prädiktorvariablen. Die Quasi-Maximum Likelihood (Quasi-ML) Methode wird deshalb zur Aufdeckung der nicht-linearen Beziehungen zwischen Zufriedenheit und Loyalität verwendet und mit der multivariaten Dummymodierungs-Regressionslösung verglichen. Es können zwei signifikant negative quadratische Terme für die nicht-lineare Beziehung zwischen Zufriedenheit mit der Marke und der anschließenden Kundenbetreuung gefunden werden. Passend zur *prospect theory* bestätigt dies die Aussage eines geringeren Einflusses auf die Loyalitätssteigerung, je höher die Zufriedenheit ist. Sie empfehlen die Untersuchung von Moderatoreffekten. Wie auch bei den quadratischen Effekten, kann aktuell nur der Ansatz von Klein multiple Interaktionen simultan schätzen.

3.5. Statische Effekte

3.5.1. Zufriedenheit und Loyalität

Fullerton und Taylor (2002) beschreiben nicht-lineare Effekte der Zufriedenheit auf loyalitätsbezogene Eigenschaften. Sie grenzen Servicequalität, also die Gesamtevaluierung des Serviceproviders, von Zufriedenheit, nämlich die

affektive oder emotionale Reaktion einer kürzlich erfahrenen Interaktion mit dem Provider, ab. *Discriminant validity* Studien können dies bestätigen.

Es entstehen Auswirkungen auf den Zufriedenheitseffekt, wenn dieser als Mediator ausgerichtet wird. Die gesamte Servicequalitätseinstellung verliert an Relevanz als Prädiktor für die Loyalitätsabsichten, wenn die Rolle des Mediators Zufriedenheit steigt. Unter völliger Mediation werden die Konsumentenabsichten rein durch die affektive Reaktion auf die jüngste Interaktion beeinflusst. Sie zeigen aber später, dass die Korrelationsmatrix der latenten Variablen hohe signifikante Korrelationen zwischen Servicequalität, Zufriedenheit und allen abhängigen Variablen aufweist. Der nicht-lineare Anstieg der Beziehung zwischen Servicequalität und positiven Absichten ist größer, wenn die Zufriedenheit höher ist bzw. niedriger bei unzufriedenen Interaktionen. Hiermit ist der *slope* der Beziehung zwischen abhängiger und unabhängiger Variable gemeint, welcher sich mit dem Level der unabhängigen Variable ändert. Diese Eigenschaft wurde oftmals unter dem Begriff *delight* erwähnt. So variiert der *slope* der Zufriedenheit-Wiederkaufabsicht-Beziehung mit dem Level der erfahrenen Zufriedenheit des Konsumenten. Nicht-lineare Zusammenhänge in Regressionen mit zusätzlichen unabhängigen Variablen, nämlich den quadrierten und dem Kubik der Originalantworten, können gefunden werden und bestätigen die Existenz nicht-linearer Beziehungen zwischen Zufriedenheit und ihren Auswirkungen. Diese nicht-linearen Beziehungen haben einen Wendepunkt auf höheren Levels der Zufriedenheit auf dem Spektrum der Erwartungserfüllung, was durch die Signifikanz des quadrierten Zufriedenheitsterms gezeigt wird. Das Vorzeichen des Terms zeigt in die gleiche Richtung wie der einfache Zufriedenheitseffekt. Dies deutet auf einen steigenden Effekt der Zufriedenheit auf die Wiederkaufabsicht, Mundpropaganda und Zahlungsbereitschaft in höheren Zufriedenheitslagen hin. So empfehlen Fullerton und Taylor (2002) Zufriedenheitslevels von ‚zufrieden‘ auf ‚sehr zufrieden‘ oder *delight* anzuheben, da der Effekt auf die Loyalität zunehmend stärker als linear ist. Die Anstrengungen dafür müssen jedoch mit den Gewinnen abgewogen werden.

3.5.2. Performance, Zufriedenheit und zukünftiges Verhalten im Zeitablauf

Mittal et al. (1999) untersuchen die Veränderung der Auswirkungen einzelner Attributsgewichte und den Zufriedenheitseinfluss auf die zukünftigen Absichten über einen direkten Effekt bzw. über den eines Mediators. Es wird davon ausgegangen, dass Kunden ihr Urteilskonstrukt nicht neu definieren, sondern updaten. Dies kommt dem *expectation-disconfirmation*-Paradigma gleich,

welches auf der *adaptation-level theory* beruht. Zusätzlich gehen sie auf den asymmetrischen Kreuzeffekt zwischen Produktzufriedenheit und Servicezufriedenheit ein, welcher sich über die Zeit hinweg umkehrt.

Die hier relevante Verbindung betrifft die Zufriedenheit zum Zeitpunkt zwei, welcher eine Beeinflussung der Gesamtzufriedenheit durch die Attributsperformance zum Zeitpunkt eins unterstellt wird. Die Zufriedenheit zum Zeitpunkt zwei, wird durch die Attributsperformance zum Zeitpunkt zwei beeinflusst, als auch die Gesamtzufriedenheit zum Zeitpunkt eins. Zeitlich kann festgestellt werden, dass sich die Gewichte der Attributsperformance auf die Zufriedenheit ändern. Dazu wird ein Modell mit Einschränkungen, welche die Gewichte gleichsetzt, mit einem frei geschätzten Modell über einen χ^2 -Vergleich verglichen.

Drei weitere Modelle werden verglichen. Im ersten hängen die Absichten zum Zeitpunkt zwei lediglich von der Zufriedenheit zum Zeitpunkt eins ab. Im Vergleich zum Gesamtmodell ist dies schlechter, darf jedoch nicht über den χ^2 -Vergleich verglichen werden. Im zweiten Modell, *dual-mediation*, führen sowohl die Zufriedenheit zum Zeitpunkt zwei, als auch die Absichten zum Zeitpunkt eins zur Beeinflussung des Zusammenhangs zwischen der Zufriedenheit zum Zeitpunkt eins und den Absichten zum Zeitpunkt zwei. Der Vergleich wird über das *Akaike Information Criterion* (AIC), *Schwartz's Bayesian Criterion* (SBC) und *Root Mean Squared Residual* (RMSR) durchgeführt. Der indirekte Effekt zwischen Zufriedenheit zum Zeitpunkt eins und Absicht zum Zeitpunkt zwei wird bestätigt. Im dritten Modell haben die Absichten zum Zeitpunkt eins einen direkten Einfluss auf die Absichten zum Zeitpunkt zwei, als auch die Zufriedenheit zum Zeitpunkt zwei. Zufriedenheit zum Zeitpunkt eins wird nicht berücksichtigt und hat keinen Einfluss auf die Zufriedenheit zum Zeitpunkt zwei, jedoch haben die Absichten zum Zeitpunkt eins Auswirkung auf die Zufriedenheit zum Zeitpunkt zwei. Da ein χ^2 -basierter Vergleich der nicht genesteten/verschachtelten Modelle wiederum nicht möglich ist, wird AIC und SBC herangezogen. Modell zwei wird mit dem niedrigeren Wert Modell drei vorgezogen. Da die Verteilung dieser Messungen nicht bekannt ist, kann die Signifikanz nicht getestet werden.

Es wird ebenfalls auf die Forschungslücke der longitudinalen Untersuchung der Nichtlinearitäten und der Asymmetrien, nämlich die stärkere Beeinflussung der schlecht bewerteten Attributsperformance und die schwächere Beeinflussung der gut bewerteten, hingewiesen.

3.6. Longitudinale Analysen

3.6.1. Longitudinale Effekte

Bolton und Drew (1991) modellieren den Effekt der Serviceveränderung auf die Evaluierung der Servicequalität. Serviceveränderungen haben einen starken Einfluss auf die Evaluierung der Servicequalität über den Effekt der Empfindung der aktuellen Performance und Diskonfirmation. Direkt nach der Serviceänderung ist der Effekt der Diskonfirmation größer. Der Effekt der vorangegangenen Einstellung ist kleiner, verglichen mit einer zeitlich vorangegangenen Periode. Es wird zwischen Zufriedenheit und Einstellung unterschieden, wobei erstere eher von einer speziellen Transaktion abhängt und letztere eine globale Evaluierung des Produktes/Service darstellt.

Satisfaction/Dissatisfaction mit einem Service hängt von der derzeitigen Empfindung, der vorangehenden Empfindung, der Erwartung von Performance und der Empfindung der Diskrepanz zwischen den beiden ab (Formel 4).

$$CS/D_t = f(DISCONFIRM_t, PERFORM_t, EXPECT_{t-1})$$

Formel 4: Customer Satisfaction/Dissatisfaction (Bolton und Drew, 1991)

Die Einstellung ist abhängig von der vorangegangenen Einstellung und des Mediators Satisfaction/Dissatisfaction (Formel 5).

$$ATTITUDE_t = g(CS/D_t, ATTITUDE_{t-1})$$

Formel 5: Attitude 1 (Bolton und Drew, 1991)

Im generellen Modell, hängt die Einstellung von der vorangegangenen Einstellung, modifiziert durch die Empfindung der aktuellen Performance, der vorangegangenen Erwartungen bezüglich Performance und der Diskrepanz zwischen Erwartung und vorangegangener Empfindung ab (Formel 6).

$$ATTITUDE_t = h(DISCONFIRM_t, PERFORM_t, EXPECT_{t-1}, ATTITUDE_{t-1})$$

Formel 6: Attitude 2 (Bolton und Drew, 1991)

Die Modellierung der Einstellung wird abgeändert und hängt von der vorangegangenen Einstellung, modifiziert durch die Empfindung der aktuellen Performance und Disconfirmation ab. Die Erwartungen aus Formel 6 werden gestrichen (Formel 7).

$$ATTITUDE_t = k(DISCONFIRM_t, PERFORM_t, ATTITUDE_{t-1})$$

Formel 7: Attitude 3 (Bolton und Drew, 1991)

Gemessen wird zu drei Zeitpunkten. Sechs Monate vor Veränderung, ein Monat danach und wiederum sechs Monate später. Die Qualität wird in Formel 8 beschrieben. LOCAL ist dabei die gesamte Empfindung der aktuellen Performance. GTE-ONLY bestimmt, ob die Person jemals in einem Gebiet gewohnt hat, welches nicht durch das Service von GTE abgedeckt wurde.

$$QUALITY_t = b_0 + d_t QUALITY_{t-1} + b_1 CHANGE_t + b_2 GTE_ONLY_t + b_3 LOCAL_t + e_t$$

Formel 8: Qualität (Bolton und Drew, 1991)

Formel 9 zeigt den Link zwischen Serviceveränderung, aktueller Performance und gesamter Servicequalität.

$$LOCAL_t = c_0 + c_1 STATIC_t + c_2 DIAL_t + c_3 CONNECT_t + c_4 CUTOFF_t + e_t$$

Formel 9: Aktuelle Performance (Bolton und Drew, 1991)

Um die statischen Empfindungen zu messen wird Formel 10 erstellt, wobei ein Koeffizient hinzugefügt wird, welcher angibt, ob sich die Person in der Kontrollgruppe oder in der Testgruppe mit den Veränderungen befindet.

$$STAT_t = a + b STAT_{t-1} + c TEST$$

Formel 10: Statische Empfindung (Bolton und Drew, 1991)

Die Empfindung der Veränderung wird in Formel 11 modelliert.

$$CHANGE_t = a + b CHANGE_{t-1} + c TEST$$

Formel 11: Veränderung (Bolton und Drew, 1991)

Die Resultate zeigen, dass eine positive Disconfirmationserfahrung einen positiven Effekt auf die Einstellung hat. Dieser Effekt, welcher über die Zufriedenheit läuft, sinkt über die Zeit. Die Empfindung der aktuellen Performance hat den größten Einfluss auf die Zufriedenheit und die Einstellung. Dieser ist größer als der Diskonfirmationseinfluss. Die aktuelle Einstellung wird durch die vorangegangene Einstellung beeinflusst. Die Beurteilung der Servicequalität hängt von der Serviceveränderung ab. Die Servicequalität ist relativ stabil, verändert sich langsam und ist daher aus langfristiger Sicht relevanter. Die Einstellungen beinhalten eine große Übertragungskomponente und resultieren nicht in einer sofortigen Veränderung. Die Einstellung wird stark

durch die aktuelle Performance bestimmt und weniger durch die Diskonfirmation. Die aktuelle Einstellung hängt stark von der vorangehenden Einstellung ab, steigt aber über die Zeit. Die Einstellung hängt mehr von den Empfindungen der aktuellen Performance und Diskonfirmation während der aktuellen Servieveränderung ab, als von Perioden ohne Veränderung. Da die Empfindungen der aktuellen Performance und Diskonfirmation übertragen werden, ist es schwer, Verbesserungen, die auf die Empfindungen wirken, zu identifizieren.

3.6.2. Zufriedenheit und Profit

Bernhardt et al. (2000) analysieren Zufriedenheit und Profitabilität über die Zeit. Sie untersuchen den Link zwischen Zufriedenheit und Performance mit Hilfe von Zeitreihendaten. Es zeigt sich eine positive signifikante Beziehung zwischen der Kundenzufriedenheit und der Mitarbeiterzufriedenheit, aber kein signifikanter Zusammenhang zwischen Kundenzufriedenheit und Performance in der *cross-sectional*-Analyse. Die Zeitreihendaten zeigen eine positiv signifikante Beziehung zwischen der Veränderung in der Kundenzufriedenheit und Veränderung in der Performance. Der Einfluss der Kundenzufriedenheit auf den Profit ist kurzfristig gegeben, aber erst langfristig signifikant positiv. So wird ein schwacher Einfluss der Zufriedenheit auf Profitmerkmale wie Verkäufe, Konsumentenanzahl und Profitabilität unterstellt. Dieser tritt aufgrund anderer Einflüsse in den Hintergrund. Es besteht jedoch ein positiver Zusammenhang zwischen Kundenzufriedenheit und den Wiederholungsabsichten sowie der positiven Mundpropaganda. Zwischen Mitarbeiterzufriedenheit und Profit wird kein signifikanter Zusammenhang angenommen, zwischen Mitarbeiterzufriedenheit und Kundenzufriedenheit ein positiver. Langfristig betrachtet wird eine Beziehung zwischen Kundenzufriedenheit und Performance angenommen, welche jedoch kurzfristig nicht zu tragen kommt, ähnlich wie zwischen Mitarbeiterzufriedenheit und Profit. Gemessen wurde die Zufriedenheit über zwölf Monate, wobei vier Paare an Messzeitpunkten gebildet wurden, um kurzfristige Fluktuationen auszuräumen. Es besteht ein positiver Zusammenhang zwischen den Serviceattributen und der Kundenzufriedenheit, keine Beziehung zwischen Kundenzufriedenheit und Profit und zwischen Mitarbeiterzufriedenheit und Profit, eine positive Beziehung zwischen Kundenzufriedenheit und den Verhaltensabsichten, sowie eine positive Beziehung zwischen Kundenzufriedenheit und Mitarbeiterzufriedenheit in den *cross-sectional*-Analysen. In den Zeitreihenanalysen zeigt sich eine positive Beziehung zwischen der sich verändernden Kundenzufriedenheit in den vorangegangenen Zeitpunkten und den sich verändernden aktuellen Profiten/Verkäufen.

Bernhardt et al. (2000) merken an, dass aufgrund der Schwierigkeit zu longitudinalen oder Zeitreihendaten zu kommen, in der Marketingforschung weitgehend *cross-sectional*-Daten verwendet werden. Sie weisen jedoch auf die Notwendigkeit hin, die Zeitverzögerung zwischen steigender Kundenzufriedenheit und steigenden Verkäufen zu analysieren. Auch *multi-item*-Konstrukte sollten erhoben werden, um analytisch bessere Methoden verwenden und die Wichtigkeit von Zufriedenheit besser zum Ausdruck bringen zu können.

3.6.3. Zufriedenheit, Wichtigkeit und Verhalten

Mittal et al. (2001) untersuchen die Veränderung des Einflusses einzelner Attribute auf die Gesamtzufriedenheit. Auch der sich wandelnde Einfluss der Gesamtzufriedenheit auf das spätere Verhalten wird untersucht. Je nachdem wie lange eine Person bereits Kontakt zu einem Unternehmen hat, wird auf unterschiedliche Attribute Wert gelegt. Dazu werden zwei Studien im Kreditkartenbereich durchgeführt. In der ersten werden Gesamtzufriedenheit und Attributperformance gemessen. Je nachdem wie lange eine Person bereits die Kreditkarte besitzt, ist der Einfluss stärker oder schwächer. Weiters wird untersucht, ob die Gewichte über die Zeit gleich bleiben, oder variieren, und ob sich diese aufgrund der Länge des Kreditkartenbesitzes unterscheiden. Über eine Regressionsanalyse aus Formel 12 werden signifikant steigende und sinkende, als auch gleich bleibende und nicht signifikante Veränderungen der Attributseinflussgewichte gefunden.

$$SAT = \beta_0(\text{Intercept}) + \beta_1(\text{INT}) + \beta_2(\text{STM}) + \beta_3(\text{BEN}) + \beta_4(\text{SRV}) + \beta_5(\text{CRE}) + \beta_6(\text{TIME}) + \beta_7(\text{INT} * \text{TIME}) + \beta_8(\text{STM} * \text{TIME}) + \beta_9(\text{BEN} * \text{TIME}) + \beta_{10}(\text{SRV} * \text{TIME}) + \beta_{11}(\text{CRE} * \text{TIME})$$

Formel 12: Veränderungsmessung in der Zufriedenheit

Die Variabilität einer Attributperformance lenkt zusätzlich die Aufmerksamkeit auf sich und so wird der Einfluss stärker. Umgekehrt haben stabile Attribute sinkenden Einfluss auf die Gesamtzufriedenheit. Dazu wird in einer zweiten Studie die Evaluierung eines Business-Marketing-Kurses herangezogen. Die Beeinflussung der Wiederbesuchsabsicht durch die Attributperformance und die Gesamtzufriedenheit wird untersucht. Die Gesamtzufriedenheit und Wiederbesuchsabsicht ändern sich im Laufe der Zeit, als auch die Beeinflussung der Wiederbesuchshäufigkeit durch die Attribute und die Gesamtzufriedenheit. Um den Zusammenhang zwischen Attributvariabilität und Stärke der Beeinflussung der Attribute zu messen, wird die Korrelation zwischen Attributseinfluss und Standardabweichung der Attributperformance herangezogen. Durch die Variabilität variiert die Stärke des Einflusses. Manche Attribute weisen diesen Zusammenhang sowohl in Bezug auf die Wiederbesuchsabsicht, als auch in

Bezug auf die Gesamtzufriedenheit auf, andere nur für eine der beiden. Dazu ist kritisch anzumerken, dass die Verwendung von Standardabweichungen zu einer schwachen Darstellung der Variabilität führt. Bewerten alle Kursbesucher in einer Periode die Performance gleich schlecht und in der nächsten Periode gleich gut, gibt es keine Variabilität. So wird hier die Standardabweichung der Veränderung der Abweichungen empfohlen. Die Differenz der individuellen Bewertung der Performance wird für jede Person berechnet und daraus die Standardabweichung gezogen. Durch Vergleich mit dem Einflussgewicht sollten realistischere Ergebnisse erzielt werden.

3.6.4. Longitudinale Messung der Servicequalität

O'Neill (2003) untersucht die Veränderung der wahrgenommenen Servicequalität evaluiert durch Studenten. Es wird von sich verändernden Wahrnehmungen und Erwartungen ausgegangen und die Idee stärkerer Diskriminierung und höherer Ansprüche wird verfolgt. Es wurden sowohl Erwartung als auch Wahrnehmung gemessen, um festzustellen ob die Erwartung erfüllt oder nicht erfüllt wurden. Weiters wurden die Wichtigkeiten der adaptierten SERVQUAL-Dimensionen beurteilt. Bezüglich der Dimensionalität des SERVQUAL-Modells konnte festgestellt werden, dass zum Zeitpunkt eins die Beurteilungen der Eindimensionalität folgt, zum Zeitpunkt zwei sich die Struktur aus mehreren Dimensionen zusammensetzt. Dies dient als Indiz zur Beurteilungsveränderung. Durch Aufaddierung der Beurteilungen können unterschiedliche Abnahmen der durchschnittlichen Bewertung für einzelne Kriterien festgestellt werden. Dies geht konform mit dem Ansatz, dass frühere Wahrnehmungen von Serviceleistungen durch veränderte Erwartungen schlechter dargestellt werden. Durch Vergleiche der Beurteilungssprünge können signifikant unterschiedlich starke Veränderungen der drei aufgefundenen Dimensionen entdeckt werden. Aufgrund der Untersuchung zu lediglich zwei Zeitpunkten ist die Veränderung zur Linearität gezwungen.

3.6.5. Kundenzufriedenheit und Treue

Cooil et al. (2007) bringen ein Beispiel für den Einfluss der Kundenzufriedenheitsveränderung auf die Veränderung der Ausgabenanteile in der Kategorie *share of wallet*. Dies ist der Prozentanteil des Geldes, den ein Konsument in einer Anlagemöglichkeit in einer Bank deponiert (Einlagen, Kredite, Investments). Sie werden als Prozent aller Finanzinstitutionen, bei welchen der Konsument Kunde ist, angegeben. Diese Effekte wurden bis dato nur in *cross-sectional*-Studien untersucht. Es wird eine positive Beziehung zwischen der Kundenzufriedenheitsveränderung und der zeitgleichen Veränderung des Geldanteiles unterstellt. Weiters soll sie von der Basis-

zufriedenheit abhängen und über die Segmente variieren. Moderatoren dieser Beziehung sind das Alter (positiv), das Einkommen (negativ), die Ausbildung (negativ), die Expertise (negativ) und die Länge der Beziehung (negativ).

Um den Anteil vorherzusagen wird eine *two-level latent class regression* verwendet. *Random effects* auf Haushaltsebene werden erlaubt. Diese sind unabhängig zwischen den Haushalten, sobald der *random effect* eingeführt wird. Die Wahrscheinlichkeit einer latenten Klasse anzugehören ist eine logistische Funktion aus Kovariaten, welche die *random effects*, mit Mittelwert null und einer Varianz entsprechend der latenten Klasse, beinhalten. Geschätzt wird mit Maximum Likelihood. Zur Modellauswahl wird das BIC minimiert, welches sich theoretisch und empirisch für viele Analysen, so auch für die latente Klassenanalyse bewährt hat. Es wird jedoch angemerkt, dass die stufenweise Prozedur nicht immer zum globalen Optimum führt. Das optimale Set an Kovariaten und Prädiktoren hängt von der Anzahl der latenten Klassen ab. Es wird vorgeschlagen, für jede plausible Anzahl an Klassen das beste Modell zu bestimmen und daraus das Beste zu wählen. Verwendet wird eine rückwärts stufenweise Eliminierung von Prädiktoren und Vereinfachungen über die Klassen, wobei die am wenigsten signifikanten Prädiktoren entfernt werden oder mit klassenunabhängigen Koeffizienten geschätzt werden, wenn sich diese über die Klassen nicht als unterschiedlich herausstellen. In jeder Stufe wird die optimale Klassenanzahl bestimmt, bis das BIC keine Verbesserung mehr zeigt. Bezüglich Modellfit wird erwähnt, dass Rückschlüsse linearer latenter Klassenmodelle unter relativ allgemeinen Bedingungen asymptotisch gültig sind und beträchtliche Abweichungen von der Normalverteilung erlauben. Sie prüfen das Modell mit diversen Residuendiagnostiken, Multikollinearität, dem geschätzten Varianzinflationsfaktor (VIF) und dem Effekt der Extreme mit der Cook-Distanz. Das beste Modell wird durch die in Prozent angegebene Fehlklassifikation identifiziert. Weiters wird die proportionale Reduktion der *entropy* der Klassifikation in Prozent, wenn die posterioren Wahrscheinlichkeiten des Modells anstatt der marginalen Wahrscheinlichkeiten jeder Klasse verwendet werden, evaluiert. Die Profile werden als abhängige Variable der latenten Klassen, durch die Veränderungen des Gesamtgeldanteiles, der zeitgleichen Veränderung der Zufriedenheit und anderen Variablen in der ersten Periode, beschrieben.

Der Effekt der Zufriedenheitsveränderung auf den Geldanteil ist positiv signifikant und variiert über die latenten Klassen und das Basiszufriedenheitslevel. Von den Moderatoreffekten sind Einkommen und Beziehungsdauer signifikant negativ. Zusammengefasst stehen Zufriedenheitsveränderungen

positiv und nicht-linear mit dem Geldanteil über die Zeit in Beziehung, wobei großteils das Zufriedenheitslevel und die bedingten Perzentile der Zufriedenheit diese Auswirkungen auslösen. *Cross-sectional*-Analysen können die Rolle der Basiszufriedenheit und Zufriedenheitsveränderungen auf den Geldanteil nicht untersuchen. Es wird auch darauf hingewiesen, dass nicht alle Kunden Homogenität dieser Beziehung aufweisen und somit wiederum Segmente notwendig sind.

3.6.6. Veränderung der Zufriedenheit

Allan et al. (2009) bringen ein Beispiel der Zufriedenheitsveränderung von Patienten bezüglich der angewandten Praktiken über eine Zeitspanne von zehn Jahren. Als Methode wurde *logistic generalized estimating equations* (LGEE) verwendet, welche Daten in diskreten Clustern analysiert. Longitudinale Daten gibt es zum Gesamtzufriedenheitslevel, wobei die Antwortvariabilität sehr schlecht ist, da großteils nur zufriedene Antworten gegeben werden. Weiters werden zwölf Attribute dichotomisiert, je nachdem ob eine Person mit allen zwölf Teilattributen zufrieden war oder zumindest mit einem nicht. Durch die größere Variabilität werden bessere Ergebnisse in Aussicht gestellt. Die Analyse mit dem Indikator, welcher mehrere Variable beinhaltet, zeigt keinen Zusammenhang zur Zeitvariablen, die der Gesamtzufriedenheit schon. Jedoch ist dieser so klein, dass dieser als vernachlässigbar beurteilt wird. Auch anhand der Sequenz der Befragung der durchgeführten Praktiken wurde die Zufriedenheit analysiert. Die Gesamtzufriedenheit zeigt keine Veränderung, der mehrere Variablen beinhaltende Indikator schon, jedoch bei der multivariablen Analyse nicht mehr. Schlussfolgernd kann zeitlich keine merkwürdige Veränderung gefunden werden.

3.7. Empirische Überprüfung der statischen Konzepte

Aus den bis dato vorgestellten Konzepten werden im Empirieteil einige ausgewählt und empirisch anhand der Kinozufriedenheitsdaten überprüft. Zuerst wird die Importance Performance Analyse (IPA) nachgerechnet und visualisiert. Probleme dieses Modelles liegen in der direkt abgefragten und somit subjektiv beurteilten Wichtigkeit und auch der Zufriedenheit. Daraus ergeben sich positiv schiefelastige Skalen, welche die tatsächliche Streubreite der Heterogenität der Stichprobe nach unten verzerren und somit verfälschen. Zusätzlich können unter Verwendung nur weniger Items einzelne Ausreißer an Items das Ergebnis ebenfalls verzerren. Durch eine extreme Positionierung eines speziellen Items wird der Gesamtmittelwert der Wichtigkeiten bzw. Zufriedenheiten nach oben bzw. unten verschoben und kann zu einer veränderten Interpretation einzelner Items aufgrund von Quadrantenverschiebungen führen.

Dieses Phänomen kommt deutlich im ebenfalls empirisch erstellten Vavras-Importance-Grid zum Tragen. Aufgrund einer notwendigen Vorselektion der Items werden dafür maximal fünf Items verwendet und es ist erkennbar, dass ein Item, nämlich die Zufriedenheit mit dem Film, den Mittelwert deutlich nach oben verschiebt. Hätte man dieses Item nicht gemessen, so wären die restlichen Attribute anders zu interpretieren. Ein weiterer statistischer Missbrauch des Vavras-Importance-Grids liegt in der Behandlung geordnet kategorialer, also ordinaler Daten, als wäre diese kontinuierlich und somit intervallskaliert. Aufgrund des eben erwähnten typischen sozialempirischen Problems schiefverteilter Skalen kann davon jedoch nicht ausgegangen werden.

Auch die Dummy-Regression wird später anhand des Kinodatensatzes durchgeführt. Auch dieser Ansatz berücksichtigt nicht die tatsächliche Skalierung der beobachteten Items. Im später folgenden Mehrebenenkapitel wird diesem Umstand Rechnung getragen und auch im Kapitel der Markovketten Modelle wird die falsche Verwendung der Skalen bereinigt. Die Ergebnisse der Dummy-Regression bringen deutlich zum Vorschein, dass eine absichtliche Suche nach *excitement*- bzw. *basic*-Faktoren notwendig ist um im Zuge der Dummy-Regression auch tatsächlich Items in allen Kategorien aufzufinden. Herkömmliche Items können meist sowohl zur Zufriedenheit als auch zur Unzufriedenheit führen.

Der vielfach erwähnte Zusammenhang zwischen Wichtigkeit und Zufriedenheit darf ebenfalls nicht außer Acht gelassen werden. In der vorliegenden Arbeit wird auf diesen deshalb anschaulich eingegangen. Da die Wichtigkeit jedoch nicht in jeder Befragungswelle abgefragt wurde, wird auf den longitudinalen Zusammenhang hier nicht detailliert eingegangen, sollte jedoch helfen Zufriedenheitseigenschaften interpretierbar und begründbar zu machen.

Die letzte statische Analyse stellt einen Methodenvorschlag dar, nämlich Paarvergleichsmodelle. Relative Wichtigkeiten werden hier nicht direkt abgefragt, sondern indirekt über die jeweiligen Vergleiche zwischen den Items. Diese Lösung verbessert die meist innewohnende Verzerrung subjektiver Wichtigkeitseinschätzungen. Für diese Methode wird auch eine Lösung über die Zeit vorgestellt, welche den Übergang zu den longitudinalen Modellen darstellt. Im vorangegangenen Kapitel wurden kurz literaturbasierte longitudinale Lösungen vorgestellt, welche aufgrund der Verwendung meist simpel gebauter Regressionsmodelle jedoch nur eingeschränkten Interpretations- und Ergebnisspielraum bieten. Der Hauptteil der empirischen Analysen ist diesem

Problem gewidmet und stellt eine Vielzahl an Lösungsvorschlägen zur Analyse longitudinaler Zufriedenheitsveränderungen vor.

4. Empirische Methodenüberprüfung und -entwicklung

Im empirischen Teil werden zuerst in der Literatur aufgefundene Methoden und Konzepte der Zufriedenheitsforschung anhand des Kinopanelns untersucht. Im Anschluss werden neuartige Methoden entfernter Forschungsgebiete zur Aufarbeitung der besprochenen Forschungslücken verwendet. Der Aufbau der Unterkapitel untergliedert sich jeweils in die Gliederungsschritte Ziele, verwendete Variablen, (verwendete Programme) und Erklärung des Modells.

4.1. Deskriptive Statistik

Befragt wurde über zwölf Monate. Insgesamt haben 5.664 Personen zumindest einen Fragebogen online ausgefüllt. 203 Respondenten waren nie im Kino und haben die relevanten Zufriedenheitsfragen deshalb nie beantwortet. Die endgültige Anzahl an Respondenten mit Zufriedenheitsinformation liegt somit bei 5.461. Das mittlere Alter der 3.213 weiblichen und 2.153 männlichen Personen liegt bei 27,52 Jahren. 4.446 Respondenten kommen aus Wien und der Rest aus Gegenden rund um Wien. Der Datensatz umfasst 485 Schüler, 51 Lehrlinge, 1.899 Studenten, 45 Präsenz- bzw. Zivildienstler, 1.955 Unselbstständige, 267 Selbstständige, 263 Beamte, 67 Pensionisten und 334 sonstige. Die meisten Personen haben drei bis sieben Kinobesuche in den zwölf Monaten unternommen, nämlich 203 keinen, 436 einen, 495 zwei, 580 drei, 561 vier, 545 fünf, 559 sechs, 547 sieben, 473 acht, 453 neun, 375 zehn, 274 elf und 163 zwölf Besuche. Insgesamt wurden 31.899 Kinobesuche bewertet, was einer durchschnittlichen Anzahl von 2.658,25 Besuchen pro Monat entspricht. Die Panelmortalität war sehr gering. Die Anzahl der ausgefüllten Fragebögen betrug 2006 im September 4.318, Oktober 4.536, November 4.926, Dezember 4.813, 2007 im Jänner 4.669, Februar 4.376, März 4.274, April 4.368, Mai 3.742, Juni 4.211, Juli 4.126 und September 3.779. Die elf abgefragten Zufriedenheitsitems weisen eine variierende Beantwortungshäufigkeit über die Personen pro Zeitpunkt auf, da die Probanden nicht gezwungen wurden, sämtliche Items in Bezug auf ihre Zufriedenheit zu beantworten. Die Respondenten wurden auch nach ihrem allgemeinen Besucherverhalten gefragt, nämlich ob nur Stammkino/Lieblingskino (714), ein Stammkino/Lieblingskino und weitere verschiedene Kinos (3.095) oder unterschiedliche Kinos (1.557), besucht werden. Die meisten Kinogehrer präferieren ein Kino, besuchen aber auch andere. Es wird deshalb davon ausgegangen, dass die Konsumenten einen erwarteten Servicequalitätslevel besitzen.

4.2. Literaturbasierte Konzepte

4.2.1. Importance-Performance Analyse

Ziele:

Die elf Zufriedenheitsattribute werden anhand ihrer direkt abgefragten Wichtigkeit und Zufriedenheit eingeteilt, aus Kapitel 1.1.2.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen im Monat September 2007);

Items (#1-11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Wie wichtig sind Ihnen folgende Merkmale eines Kinos? (gemessen im Monat September 2007);

Items (#1-11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige Likert-Skala: 6 (sehr wichtig) bis 1 (nicht wichtig);

Modell:

Es werden aus den gültigen Teilzufriedenheitsbewertungen aller Kinobesuche im Monat September 2006 die Zufriedenheitsmittelwerte und die explizit befragten Wichtigkeiten der Serviceattribute berechnet. In Diagramm 30 liegt die vertikale Hauptachse beim *grand mean* der Mittelwerte aller elf Zufriedenheitsvariablen. Die horizontale Hauptachse liegt beim *grand mean* der Mittelwerte aller elf Wichtigkeitsvariablen. Auffällig ist, dass der *grand mean* der Gesamtmittelwerte der Zufriedenheitsvariablen, 1,878333, als auch der *grand mean* der Gesamtmittelwerte der Wichtigkeitsvariablen, 1,875, nahe beieinander liegen. Sie weisen beide sehr positive Zufriedenheitswerte bzw. hohe Wichtigkeitswerte auf.

Um zu zeigen, dass die Wichtigkeits- und Zufriedenheitswerte im September 2007 unterschiedliche Aspekte darstellen, wird zusätzlich eine Pearsonkorrelation zwischen den Mittelwerten der 2x11 Variablen vorgenommen. Dabei zeigt sich ein Korrelationskoeffizient von 0,4879 bei einer Irrtumswahrscheinlichkeit von 10,7552 % und wird somit als nicht signifikant

abgelehnt. Weiter kann festgestellt werden, dass die Probanden, was die Wichtigkeiten betrifft, stärker differenzieren, verglichen zur Differenzierung in den Zufriedenheitsvariablen. Dies wird anhand der unterschiedlichen Streubreite der Punkte über die Abszisse (Zufriedenheit) bzw. Ordinate (Wichtigkeit) verdeutlicht. Buffet, Servicepersonal und Erscheinungsbild/Image sind den Probanden weder wichtig, noch sind diese damit zufrieden. Sie können entweder verbessert werden, um durch Zufriedenheit und verstärkte Inanspruchnahme deren Einfluss zu steigern, oder sie werden ignoriert. In den Variablen Standort/Erreichbarkeit, Filmangebot, Technik/Bild/Ton, Komfort und Film bieten die Kinos durchschnittlich gutes Service und dies sollte beibehalten werden. Atmosphäre und Preis/Leistung ist den Probanden wichtig, aber sie sind damit nicht zufrieden. Darin steckt verstärkt Verbesserungspotential. Informationsangebot und Reservierungsmöglichkeiten sind den Probanden nicht wichtig, sie sind aber mit der Erfüllung sehr zufrieden, was auf Kapazitäts- einsparungsmöglichkeiten hindeutet.

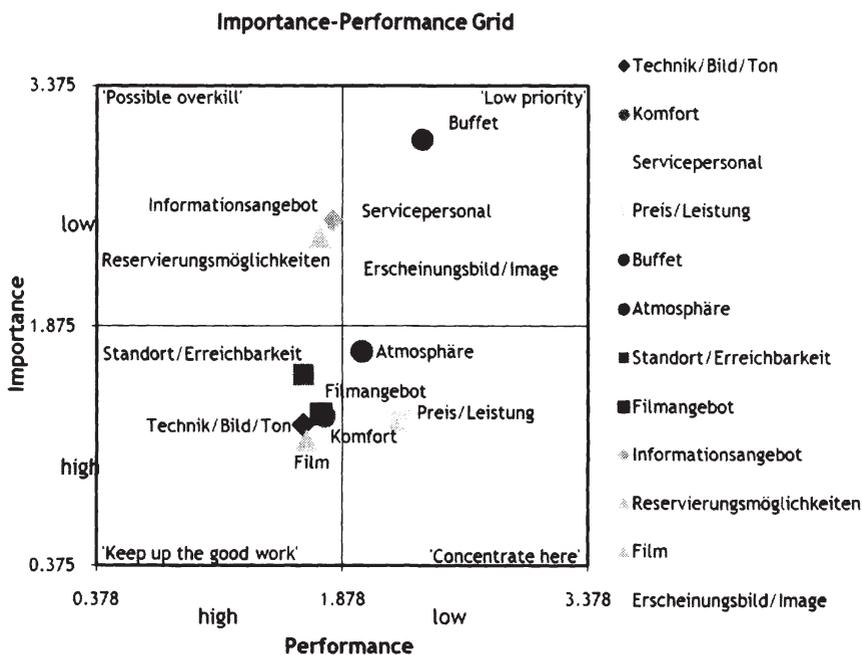


Diagramm 30: Importance Performance Grid

4.2.2. Dummy-Regression

Ziele:

Untersucht werden die Auswirkungen der Teilzufriedenheitsvariablen auf die Gesamtzufriedenheit anhand der Dummyregression, aus Kapitel 1.2.1.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den zwölf Monaten Oktober 2006 bis September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Bitte machen Sie folgende Angaben zu Ihrem Kinobesuch: Gesamtzufriedenheit (gemessen in den zwölf Monaten Oktober 2006 bis September 2007);

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Modell:

Zuerst werden die elf Zufriedenheitsvariablen dupliziert und in weiterer Folge die jeweiligen Paare in Dummyvariablen umkodiert. Für die positiven Dummyvariablen, in Tabelle 12 Spalte Dummy mit p bezeichnet, werden die Werte für ‚sehr zufrieden (6)‘ auf 1, die restlichen Werte auf 0 gesetzt. Für die negativen Dummyvariablen, in Tabelle 12 Spalte Dummy mit n bezeichnet, behalten die Werte für ‚nicht zufrieden (1)‘ ihren Wert. Die restlichen Werte werden auf 0 gesetzt. Weiters befindet sich in der Datenmatrix die Variable Gesamtzufriedenheit mit dem jeweiligen Kinobesuch. Insgesamt befinden sich in dem Datensatz Kinobesuchszufriedenheitsbewertungen zu 31.871 Besuchen. Darin kommen Probanden mehrere Male vor. Um eine Verzerrung der expliziten Wichtigkeitsstatements zu vermeiden, werden die impliziten Wichtigkeiten hier im Zuge einer multiplen Regression berechnet. Abhängige Variable ist die Gesamtzufriedenheit mit dem jeweiligen Kinobesuch. Die unabhängigen Variablen bilden die elf negativen und positiven Dummyvariablen. Ziel ist es mit Hilfe der PRC-Analyse, durch ausschließliche Regressierung der Extremwerte der unabhängigen Variablen, zwischen *excitement*- und *basic*-Faktoren zu unterscheiden. Jene Variablen, bei welchen sich in den zusammengehörigen standardisierten Regressionskoeffizienten rein positive Werte finden, werden als *excitement*-Faktoren mit einer rein positiven Auswirkung auf die Gesamtzufriedenheit ausgewiesen. Jene Variablen, bei welchen sich in den

zusammengehörigen standardisierten Regressionskoeffizienten rein negative Werte finden, werden als *basic*-Faktoren ausgewiesen, welche lediglich eine negative Auswirkung auf die Gesamtzufriedenheit ausüben. Jene Variablen, bei welchen sich in den zusammengehörigen standardisierten Regressionskoeffizienten sowohl positive als auch negative Werte finden, können sowohl positive als auch negative Auswirkung auf die Gesamtzufriedenheit ausüben. Es werden nur 12,4 % der Gesamtzufriedenheit mit den 22 Dummy-Variablen erklärt. Tabelle 12 zeigt die relevanten Ergebnisse der multiplen Regression. Im Modell wird ein Signifikanzniveau mit einer Irrtumswahrscheinlichkeit von 5% vorausgesetzt. Signifikante Attribute werden, wie auch in den folgenden Tabellen, in der Spalte ‚Signifikanz‘ mit einem Stern-(*) markiert.

Dummy	Variable	Standardisierte Regressionskoeffizienten	Fehlerwert	Signifikanz
	(Constant)		0	
p	Technik/Bild/Ton	0,100	0,000	*
p	Komfort	0,035	0,000	*
p	Servicepersonal	0,027	0,002	*
p	Preis/Leistung	0,080	0,000	*
p	Buffett	-0,010	0,227	
p	Atmosphäre	0,077	0,000	*
p	Standort/Erreichbarkeit	0,033	0,000	*
p	Filmangebot	0,049	0,000	*
p	Informationsangebot	-0,015	0,117	
p	Reservierungsmöglichkeiten	0,015	0,061	
p	Erscheinungsbild/Image	0,043	0,000	*
n	Technik/Bild/Ton	-0,052	0,000	*
n	Komfort	-0,042	0,000	*
n	Servicepersonal	-0,033	0,000	*
n	Preis/Leistung	-0,052	0,000	*
n	Buffett	-0,020	0,005	*
n	Atmosphäre	-0,051	0,000	*
n	Standort/Erreichbarkeit	-0,016	0,030	*
n	Filmangebot	-0,015	0,046	*
n	Informationsangebot	-0,001	0,851	
n	Reservierungsmöglichkeiten	-0,010	0,160	
n	Erscheinungsbild/Image	-0,034	0,000	*

Tabelle 12: Dummy-Regression 12345-6

Zufriedenheitskriterien, bei welchen sowohl die positiven als auch negativen Dummyvariablen signifikante Werte aufweisen, finden sich in den Variablen Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Atmosphäre, Standort/Erreichbarkeit, Filmangebot und Erscheinungsbild/Image. Alle Zufriedenheitsvariablen, welche in beiden Dummy-Variablen signifikante standardisierte Regressionskoeffizienten aufweisen, besitzen sowohl einen positiven als auch einen negativen Wert. Es wird weder ein *excitement*- noch ein *basic*-Faktor aufgefunden. Diagramm 31 zeigt die visualisierte Lösung.

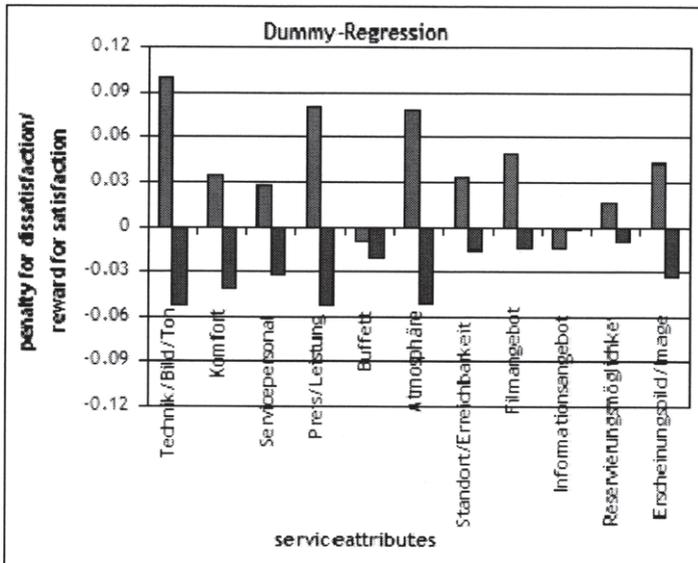


Diagramm 31: Dummy-Regression 12345-6

Die Analyse wird erneut mit einer abgeschwächten Version der Extremwert-Dummy-Berechnung durchgeführt. Die elf Zufriedenheitsvariablen werden wiederum dupliziert. In weiterer Folge werden jedoch die Dummy-Variablen im Vergleich zu obigem Beispiel unterschiedlich recodiert. Für die positiven Dummyvariablen werden die Werte für ‚sehr zufrieden (6)‘ und ‚zufrieden (5)‘ auf 1 gesetzt, die restlichen Werte auf 0. Für die negativen Dummyvariablen werden die Werte für ‚kaum zufrieden (2)‘ und ‚nicht zufrieden (1)‘ auf 1 gesetzt und die restlichen Werte auf 0.

Es werden 13 % der Gesamtzufriedenheit mit den 22 Dummy-Variablen erklärt. Zufriedenheitskriterien, bei welchen sowohl die positiven als auch negativen Dummyvariablen signifikante Werte aufweisen finden sich in den Variablen

Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Atmosphäre, Standort/Erreichbarkeit und Erscheinungsbild/Image. Diese Variablen können aus Diagramm 32 aussagekräftig interpretiert werden. Alle Zufriedenheitsvariablen, welche in beiden Dummy-Variablen signifikante standardisierte Regressionskoeffizienten aufweisen, besitzen sowohl einen positiven als auch einen negativen Wert auf. Es wird wiederum weder ein *excitement*- noch ein *basic*-Faktor nachgewiesen.

Dummy	Variable	Standardisierte Regressionskoeffizienten	Fehlerwert	Signifikanz
	(Constant)		0,000	
p	Technik/Bild/Ton	0,071	0,000	*
p	Komfort	0,056	0,000	*
p	Servicepersonal	0,044	0,000	*
p	Preis/Leistung	0,102	0,000	*
p	Buffett	0,024	0,003	*
p	Atmosphäre	0,080	0,000	*
p	Standort/Erreichbarkeit	0,028	0,000	*
p	Filmangebot	0,031	0,000	*
p	Informationsangebot	0,000	0,963	
p	Reservierungsmöglichkeiten	0,014	0,089	
p	Erscheinungsbild/Image	0,040	0,000	*
n	Technik/Bild/Ton	-0,051	0,000	*
n	Komfort	-0,031	0,000	*
n	Servicepersonal	-0,025	0,001	*
n	Preis/Leistung	-0,024	0,001	*
n	Buffett	0,001	0,883	
n	Atmosphäre	-0,040	0,000	*
n	Standort/Erreichbarkeit	-0,025	0,001	*
n	Filmangebot	-0,012	0,119	
n	Informationsangebot	-0,005	0,472	
n	Reservierungsmöglichkeiten	-0,011	0,135	
n	Erscheinungsbild/Image	-0,024	0,004	*

Tabelle 13: Dummy-Regression 1234-56

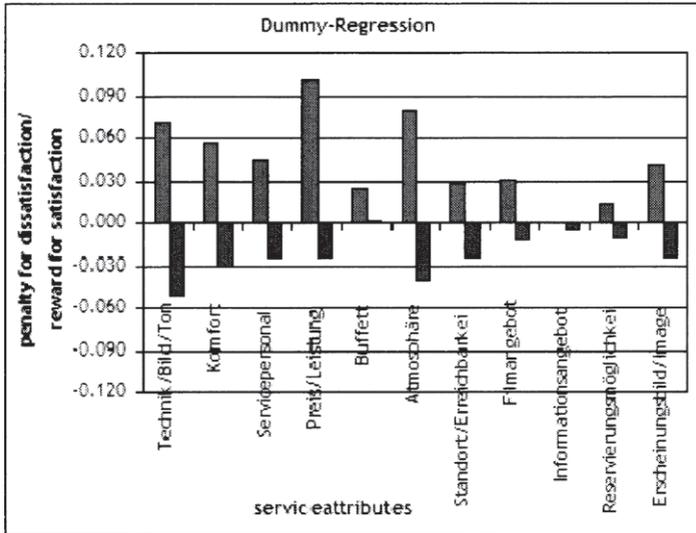


Diagramm 32: Dummy-Regression 1234-56

4.2.3. Vavras-Importance-Grid

Ziele:

Untersucht werden die Unterschiede der Wichtigkeitsbewertungen und Auswirkungen der Teilzufriedenheitskriterien auf die Gesamtzufriedenheit, aus Kapitel 1.2.1.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den Monaten November 2006 und September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffett, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Wie wichtig sind Ihnen folgende Merkmale eines Kinos? (gemessen in den Monaten November 2006 und September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffett, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr wichtig) bis 1 (nicht wichtig);

Bitte machen Sie folgende Angaben zu Ihrem Kinobesuch: Gesamtzufriedenheit (gemessen in den Monaten November 2006 und September 2007); Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Modell:

Es werden die Summen der elf Serviceattribute aus den explizit befragten Wichtigkeitsbewertungen aller Kinostandorte in den Monaten Oktober 2006 und September 2007 berechnet. Zuerst wird das Modell für Oktober 2006 gerechnet, im Anschluss für September 2007. Es folgt die Beschreibung zur Durchführung der Analyse für Oktober 2006. Zuerst wird eine multiple Regression mit der Gesamtzufriedenheit des Kinobesuches im Oktober 2006 als abhängige und den elf Zufriedenheitsbewertungen als unabhängige Variablen durchgeführt. In Tabelle 14 finden sich vier signifikante Serviceattribute, nämlich Technik/Bild/Ton, Preis/Leistung, Atmosphäre und Standort/Erreichbarkeit. 4.447 Probanden bewerteten die Wichtigkeit. Für die Zufriedenheitsbewertungen fällt die Probandenzahl geringer aus.

Variable	Standardisierte Regressionskoeffizienten	t-Werte	Signifikanz	
(Constant)		4,419	,000	
Technik/Bild/Ton	,109	4,333	,000	*
Komfort	,003	,112	,911	
Servicepersonal	,034	1,261	,207	
Preis/Leistung	,101	3,963	,000	*
Buffet	,002	,064	,949	
Atmosphäre	,112	3,653	,000	*
Standort/Erreichbarkeit	,099	4,000	,000	*
Filmangebot	,009	,323	,747	
Informationsangebot	,001	,032	,974	
Reservierungsmöglichkeiten	,039	1,592	,112	
Erscheinungsbild/Image	,050	1,616	,106	

Tabelle 14: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei)

Um Verzerrungen der nicht signifikanten Serviceleistungsattribute auszuschließen werden die vier signifikanten unabhängigen Variablen separat auf die Gesamtzufriedenheit regressiert und in Tabelle 15 dargestellt. Es werden 11,8 % der Gesamtzufriedenheit erklärt. Die standardisierten Regressionskoeffizienten stellen die Werte für die implizite Wichtigkeit dar. Diese Werte werden auf der y-Achse aufgetragen.

Variable	Standardisierte Regressionskoeffizienten	t-Werte	Signifikanz	
(Constant)		8,885	,000	
Technik/Bild/Ton	,140	7,663	,000	*
Preis/Leistung	,112	5,958	,000	*
Atmosphäre	,179	9,100	,000	*
Standort/Erreichbarkeit	,071	3,819	,000	*

Tabelle 15: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei, neu)

Für die Bestimmung der expliziten Wichtigkeiten werden jeweils die Summen aller Wichtigkeitsbewertungen der vier signifikanten Variablen aufsummiert. Im Anschluss werden diese Summen erneut aufsummiert und die Einzelsummen durch die gebildete Gesamtsumme der Wichtigkeitsbewertungen im Modell dividiert, kurz deren relativer Anteil an der Gesamtsumme ermittelt. Tabelle 16 zeigt vier Prozentwerte, welche jeweils im Vergleich der vier, die relative Wichtigkeit zu den anderen darstellt.

Variable	Summe der Wichtigkeitsbewertungen	Anteil %
Technik/Bild/Ton	6.192	23,16
Preis/Leistung	6.168	23,07
Atmosphäre	7.667	28,67
Standort/Erreichbarkeit	6.706	25,08
Gesamtsumme der signifikanten Wichtigkeitsbewertungen	26.733	

Tabelle 16: Explizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei)

Um zu zeigen, dass die impliziten und die expliziten Wichtigkeitswerte unterschiedliche Wichtigkeitsaspekte beschreiben, wird zusätzlich eine Pearson-Korrelation zwischen den Werten der vier Variablen durchgeführt. Dabei zeigt sich ein Korrelationswert von 0,534 bei einer Irrtumswahrscheinlichkeit von 46,6 % und wird als nicht signifikant abgelehnt. Die beiden Wichtigkeitsarten können inhaltlich unterschiedlich interpretiert werden.

In Diagramm 33 liegt die vertikale Hauptachse beim *grand mean* der zuvor beschriebenen Prozentsätze, welcher bei vier Variablen 25 % ergibt. Die horizontale Hauptachse liegt beim *grand mean* der standardisierten Regressionskoeffizienten. Die beiden Achsen bilden die Schnittpunkte der Übergänge zwischen explizit ‚nicht wichtig‘ und ‚sehr wichtig‘ bzw. implizit ‚nicht wichtig‘ und ‚sehr wichtig‘. Preis/Leistung wird als wichtig bewertet,

spielt aber bei der Beeinflussung der Gesamtzufriedenheit eine schwächere Rolle. Diese Serviceeigenschaft kann zwar nicht Zufriedenheit erzeugen, aber sehr wohl Unzufriedenheit. Es muss ein Mindestmaß an Qualität gewährleistet werden, um Unzufriedenheit auszuschließen. Eine Übererfüllung führt jedoch nicht zu einer Steigerung der Zufriedenheit. Umgekehrt ist Atmosphäre als unwichtig bewertet worden, hat aber einen unterbewussten Einfluss auf die Gesamtzufriedenheit. Sie wird zwar keine Unzufriedenheit erzeugen, ist aber dazu geeignet die Zufriedenheit zu steigern. Somit wird eine schlechtere Atmosphäre nicht sofort zur Unzufriedenheit der Besucher führen, eine positive Beeinflussung der Atmosphäre führt jedoch zu einer erhöhten Zufriedenheit und somit lohnt es sich, Verbesserungsbestrebungen vorzunehmen. Technik/Bild/Ton und Standort/Erreichbarkeit stellen *performance*-Faktoren dar. Diese sind sowohl zur Steigerung der Zufriedenheit als auch der Unzufriedenheit geeignet. Je nachdem wie gut die Erfüllung dieser Eigenschaften erbracht wird, desto zufriedener bzw. unzufriedener werden die Besucher sein.

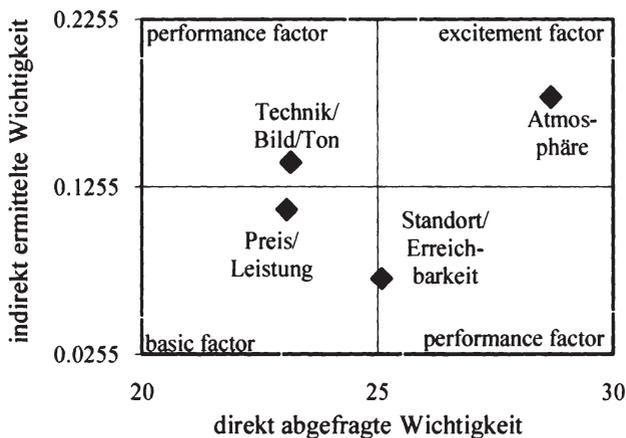


Diagramm 33: Vavras-Importance-Grid (Welle zwei)

Selbiges Modell wird für Welle zwölf im September 2008 durchgeführt, da in dieser Welle zusätzlich die Zufriedenheit und Wichtigkeit des Films abgefragt wurden. In der multiplen Regression finden sich laut Tabelle 17 fünf signifikante Serviceattribute, nämlich Technik/Bild/Ton, Preis/Leistung, Servicepersonal, Komfort und Film. 3.756 Probanden haben die Wichtigkeit bewertet. Für die Zufriedenheitsbewertungen fällt die Probandenzahl geringer aus. Um die Auswirkungen des Filmeinflusses zu visualisieren, wird das Signifikanzniveau auf eine Irrtumswahrscheinlichkeit von 6,3 % abgeschwächt und die Variable

Film in die verkürzte Version der multiplen Regression in Tabelle 18 aufgenommen. Es werden 25,5 % der Gesamtzufriedenheit erklärt. Die deutlich bessere Erklärung der Gesamtzufriedenheit im Vergleich zum vorherigen Modell wird auf die Variable Film zurückgeführt.

Variable	Standardisierte Regressionskoeffizienten	t-Werte	Signifikanz	
(Constant)		1,591	,112	
Technik/Bild/Ton	,147	4,685	,000	*
Komfort	,077	2,296	,022	*
Servicepersonal	,101	3,034	,002	*
Preis/Leistung	,071	2,227	,026	*
Buffet	,051	1,602	,109	
Atmosphäre	,009	,238	,812	
Standort/Erreichbarkeit	,001	,032	,974	
Filmangebot	,008	,227	,821	
Informationsangebot	-,047	-1,249	,212	
Reservierungsmöglichkeiten	-,017	-,567	,571	
Erscheinungsbild/Image	,073	1,861	,063	*

Tabelle 17: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwölf)

Variable	Standardisierte Regressionskoeffizienten	t-Werte	Signifikanz	
(Constant)		2,483	,013	
Technik/Bild/Ton	,123	5,074	,000	*
Komfort	,092	3,574	,000	*
Servicepersonal	,124	5,070	,000	*
Preis/Leistung	,091	3,874	,000	*
Film	,316	14,729	,000	+

Tabelle 18: Implizite Wichtigkeiten - Vavras-Importance-Grid Ergebnisse (Welle zwölf, neu)

Die Bestimmung der expliziten Wichtigkeit wird in Tabelle 19 gezeigt.

Der Pearson-Korrelationskoeffizient zwischen impliziten und expliziten Wichtigkeitswerten der fünf Variablen zeigt einen Wert von -,240 bei einer Irrtumswahrscheinlichkeit von 69,8 % und wird als nicht signifikant abgelehnt. Die beiden Wichtigkeitsarten werden wiederum inhaltlich unterschiedlich interpretiert. Diagramm 34 zeigt das Ergebnis.

Variable	Summe der Wichtigkeitsbewertungen		
Technik/Bild/Ton	4.715	*	16,67 %
Komfort	4.914	*	17,38 %
Servicepersonal	9.461	*	33,46 %
Preis/Leistung	4.810	*	17,01 %
Film	4.372	*	15,46 %
Gesamtsumme aller signifikanten Wichtigkeitsbewertungen	28.272		

Tabelle 19: Explizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwölf)

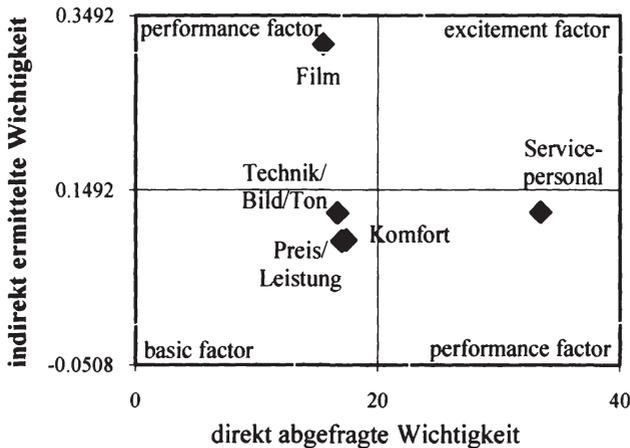


Diagramm 34: Vavras-Importance-Grid (Welle zwölf)

Preis/Leistung, Technik/Bild/Ton und Komfort werden als wichtig bewertet, spielen aber bei der Beeinflussung der Gesamtzufriedenheit eine schwächere Rolle. Sie sind zwar nicht geeignet Zufriedenheit zu erzeugen, können aber Unzufriedenheit bewirken. Für diese Serviceeigenschaften muss ein Mindestmaß an Qualität gewährleistet werden, um Unzufriedenheit auszuschließen. Eine Übererfüllung führt jedoch nicht zur Steigerung der Zufriedenheit. Film und Servicepersonal stellen *performance*-Faktoren dar. Diese sind sowohl zur Steigerung der Zufriedenheit als auch der Unzufriedenheit geeignet. Je nachdem wie gut die Erfüllung dieser Eigenschaften erbracht wird, desto zufriedener bzw. unzufriedener sind die Besucher. Der große Abstand der Variablen Film auf der implizit ermittelten Wichtigkeitsachse weist auf deren enormen Einfluss auf die Gesamtzufriedenheit hin. Für die Betrachtung der Wichtigkeiten und Unterscheidung zwischen *performance*-, *basic*- und *excitement*-Faktoren spielt

der Film eine große Rolle, da dieser die Trennung der x- und y-Achse wesentlich beeinflusst. Diagramm 34 ohne *excitement*-Faktoren scheint ein besseres Abbild der Faktoreneinteilung zu bieten. Die abgefragten Variablen stellen nämlich klassische und keine außergewöhnlichen Serviceattribute eines Kinos dar.

4.2.4. Zusammenhang zwischen Wichtigkeit und Zufriedenheit

Ziele:

Untersucht wird der Zusammenhang zwischen Zufriedenheit und Wichtigkeit, aus Kapitel 1.1.2.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen im Monat September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image, Film;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Wie wichtig sind Ihnen folgende Merkmale eines Kinos? (gemessen im Monat September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image, Film;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr wichtig) bis 1 (nicht wichtig);

Wenn Sie an Ihre Kinobesuche denken, seit Sie an der CINEMON-Studie teilnehmen, waren Sie dann:

- in den meisten Fällen im selben Kino?
- Eher in verschiedenen Kinos?

Für all jene Probanden, welche in den meisten Fällen im selben Kino waren, folgt die Frage:

Verraten Sie uns bitte, um welches Kino es sich dabei handelt. Bitte auswählen: (Drop-Down-Menü: 53 Kinostandorte zur Auswahl, gemessen im Monat März 2007);

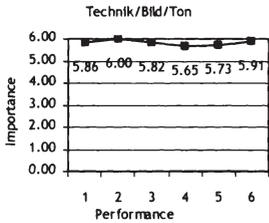
Modell:

Zuerst wird das Modell für die ‚Lieblingskinobesucher‘ (757 Probanden), im Anschluss das für die ‚Nicht-Lieblingskinobesucher‘ (4.907 Probanden) erörtert. Es werden jeweils die Probanden der konträren Besuchergruppe entfernt. Die

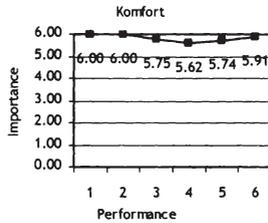
Annahme wird überprüft, ob bei sinkender Zufriedenheit die Wichtigkeit steigt, da sich die Serviceeigenschaft in den Blickpunkt des Interesses verschiebt. Bei steigender Zufriedenheit sinkt die Wichtigkeit, da die Serviceeigenschaft an Relevanz verliert, als gegeben angenommen wird und sich ein gewisses Niveau bzw. Anspruchslevel festigt. Im ersten Modell wird die Annahme für die ‚Lieblingskinobesucher‘ überprüft. Es wird davon ausgegangen, dass jenes Kino, welches in Welle sechs im März 2007 nach Aussage der Probanden in den meisten Fällen besucht wird, auch jenes ist, an welchem die Probanden in Welle zwölf im September 2008 ihre Wichtigkeitsmaßstäbe setzen. Es stellt jenes Kino dar, mit welchem der Proband andere Kinos vergleicht, da er seine Wichtigkeitsbewertungen am ehesten damit identifiziert. Jene Personen, welche auf die Zufriedenheitsfrage in Welle zwölf im September 2007 mit ‚keine Erfahrung‘ geantwortet haben, werden trotzdem in der Auswertung der Frage nach der Wichtigkeit zu diesem Serviceattribut berücksichtigt, da davon ausgegangen wird, dass die Personen zwar Erfahrung mit diesem Servicekriterium besitzen, dieses in Welle zwölf jedoch nicht genutzt und deshalb nicht bewertet haben. Zuerst werden die Zufriedenheitskategorien der einzelnen Serviceattribute aufgeteilt. Es entstehen zwölf Variable mit jeweils sechs Gruppen. Sechs Gruppen deshalb, da den Probanden eine 6-Punkt-Likert-Skala präsentiert wurde. Für die resultierenden 72 Untergruppen, zwölf Servicekriterien mal sechs Zufriedenheitsbewertungsstufen, werden die Mittelwerte der zugehörigen Wichtigkeitswerte berechnet. In den Diagrammen wird auf der x-Achse die Zufriedenheitsbewertungskategorie 1 bis 6 aufgetragen und auf der y-Achse der Mittelwert der zugehörigen Wichtigkeitsbewertungen. Der Annahme entsprechend muss der Wichtigkeitswert auf der y-Achse mit zunehmender Zufriedenheitskategorie auf der x-Achse sinken. Diagramm 35 zeigt die Visualisierungen der zwölf Serviceleistungsattribute. Unter den Diagrammen befinden sich die Probandenzahlen in den sechs Zufriedenheitskategorien in aufsteigender Reihenfolge, aus welchen die Wichtigkeitsmittelwerte gebildet wurden. Da die Verteilung der Zufriedenheitsbewertungen schief ist, da mehr Probanden ihre Zufriedenheitsempfindung auf der positiven Seite sehen, finden sich in den schlechter bewerteten Zufriedenheitskategorien weniger Probanden für die Mittelwertbildung der Wichtigkeiten. Deshalb sollten Interpretationen des Verlaufes lediglich für die Zufriedenheitsbewertungskategorien 6, 5, 4 und eventuell 3 vorgenommen werden, um aussagekräftige Schlussfolgerungen zu gewährleisten.

Alle Serviceattribute verzeichnen eine sinkende Zufriedenheit bei sinkender Wichtigkeit bzw. eine steigende Zufriedenheit bei steigender Wichtigkeit. Diese Ergebnisse stehen konträr zu den zuvor erwähnten Annahmen. Je zufriedener die

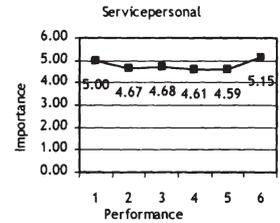
Probanden mit den einzelnen Teilzufriedenheitsattributen sind, desto wichtiger sind diese für sie. Dies zeigt, dass sich die Probanden weniger auf Serviceleistungen konzentrieren, mit welchen sie unzufrieden sind, jedoch mehr auf Serviceleistungen, mit welchen sie zufrieden sind. Dies kann möglicherweise dadurch begründet werden, dass die Probanden den unzufriedenen Serviceangeboten ausweichen und weniger Beachtung schenken. Die parallele Veränderung in der Tendenz zwischen Zufriedenheits- und Wichtigkeitsbewertungen könnte auch dadurch erklärt werden, dass ‚Lieblingskinobesucher‘ Serviceleistungen, welche sie schlechter bewerten, in ihrem eben meistbesuchten Kino akzeptieren, eben genau deshalb, weil sie diese als unwichtig empfinden. Sie wählen dieses Kino, da dort die Zufriedenheit mit jenen Serviceleistungen, welche sie als wichtig empfinden, gewährleistet wird.



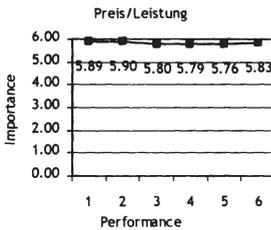
7, 7, 11, 34, 258, 434



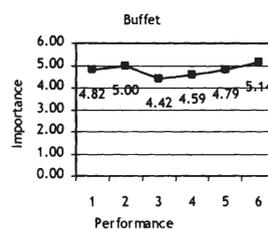
1, 2, 12, 79, 302, 357



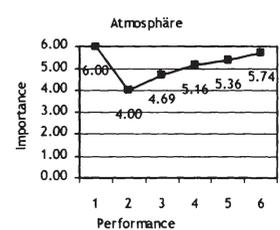
4, 3, 22, 112, 302, 247



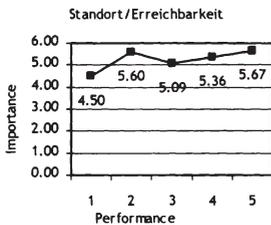
9, 21, 4, 167, 305, 204



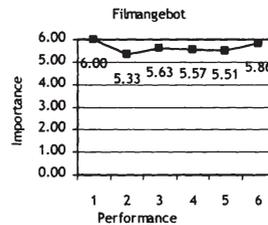
11, 12, 60, 134, 229, 133



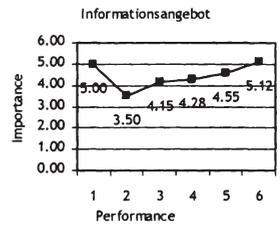
2, 3, 26, 114, 293, 314



0, 2, 5, 35, 186, 525



2, 6, 19, 83, 235, 406



1, 2, 13, 76, 269, 343

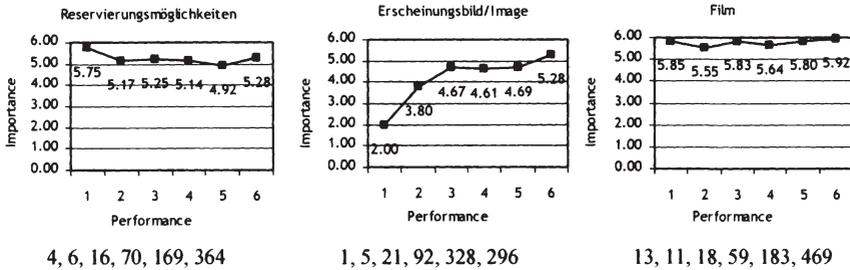
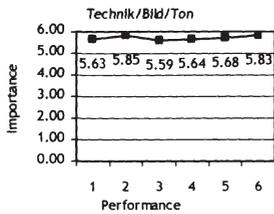


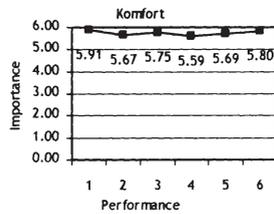
Diagramm 35: Zufriedenheit und Wichtigkeit der Lieblingskinobesucher

Um einen möglichen Unterschied zwischen ‚Lieblingskinobesuchern‘ und ‚Nicht-Lieblingskinobesuchern‘ zu überprüfen, werden all jene Probanden betrachtet, welche in Welle zwölf im September 2007 ein Kino besucht haben, welches sie in Welle sechs im März 2007 nicht als ihr meistbesuchtes Kino genannt haben. Diese Teilstichprobe ergibt das Gegenstück zur zuvor untersuchten ‚Lieblingskinobesucher‘-Teilstichprobe. In Diagramm 36 sollte sich eine abgeschwächte Tendenz zeigen, da im Vergleich zum Lieblingskino, Zufriedenheitsverstöße möglicherweise stärker geahndet werden und in den Fokus der Probanden rücken. Jene Personen, welche auf die Zufriedenheitsfrage in Welle zwölf im September 2007 mit ‚keine Erfahrung‘ geantwortet haben, werden trotzdem in der Auswertung der Frage nach der Wichtigkeit zu diesem Serviceattribut berücksichtigt, da wiederum davon ausgegangen wird, dass die Personen zwar eine allgemeine Erfahrung mit diesem Servicekriterium besitzen, jedoch dieses in Welle zwölf nicht genutzt haben und somit nicht bewerten konnten. Die Berechnung und Visualisierung erfolgt analog zu oben geschildertem Modell mit den ‚Lieblingskinobesuchern‘.

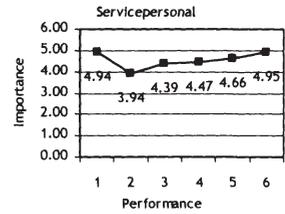
Für die Zufriedenheitskategorien fallen in der Teilstichprobe der Nicht-Lieblingskinobesucher, im Vergleich zu den Lieblingskinobesuchern, viele der Wichtigkeitsbewertungen niedriger aus. Dies kann darauf hinweisen, dass durch die über die Zeit aufgebaute Bindung zu einem Standort, nebensächlich empfundene Serviceleistungen an Wichtigkeit gewinnen. Durch die wiederholte Konfrontation wird der gleichbleibenden Erfüllung der Serviceleistungen weniger Beachtung geschenkt und die dadurch freiwerdenden Reize bieten Raum für andere Aufmerksamkeitspotentiale zusätzlicher Serviceleistungsattribute.



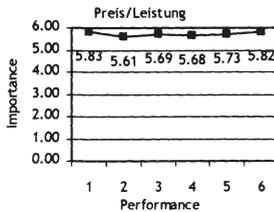
16, 13, 22, 118, 427, 660



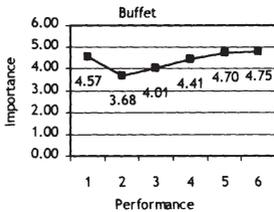
11, 9, 32, 180, 493, 534



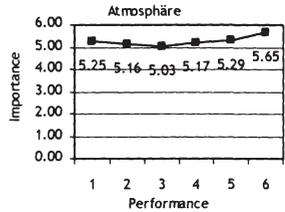
17, 16, 46, 162, 449, 400



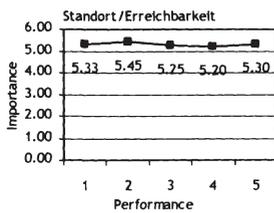
29, 36, 75, 277, 445, 358



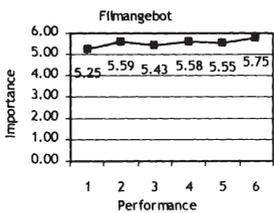
21, 22, 75, 217, 317, 204



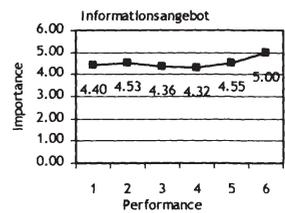
20, 25, 73, 242, 484, 414



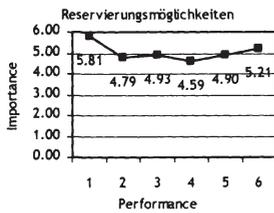
9, 22, 51, 166, 392, 620



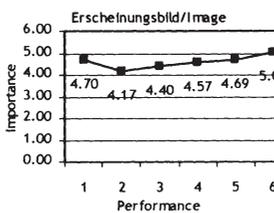
4, 17, 44, 166, 429, 554



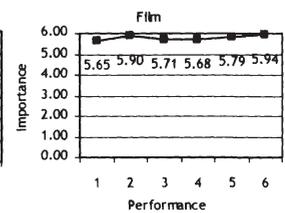
5, 17, 36, 189, 422, 429



16, 14, 29, 97, 261, 409



20, 18, 72, 226, 506, 397



23, 20, 45, 117, 309, 748

Diagramm 36: Zufriedenheit und Wichtigkeit der Nicht-Liebingskinobesucher

Auffällig sind die abweichend hohen Wichtigwerte bei niedriger Zufriedenheit in den Variablen Servicepersonal, Buffet, Reservierungsmöglichkeiten und Erscheinungsbild/Image. Die Probandenzahlen sind in dieser Kategorie mit schlechter Performance zwar sehr gering, weisen jedoch auf steigende Wichtigkeit bei extremer Enttäuschung hin. Diese Ausreißer könnten für jene Probanden stehen, welche mit einer Serviceleistung extrem unzufrieden waren

und für welche sich deshalb ein derartiges Servicequalitätsversäumnis negativ auswirkt. Bei einer extrem großen Abweichung zwischen Erwartung und Erfüllung tritt die Wichtigkeit wie im Theorieteil angesprochen in den Vordergrund. Dieses Phänomen kann für die vier genannten Serviceleistungsattribute bei der schlechtesten Bewertung gezeigt werden. Um mehr über diese vier Ausreißervariablen zu erfahren, werden deren Wichtigkeitsmittelwerte über alle sechs Zufriedenheitskategorien in Tabelle 20 berechnet.

Variable	Wichtigkeitsmittelwerte	Ausreißer
Technik/Bild/Ton	5,70	
Komfort	5,73	
Servicepersonal	4,56	*
Preis/Leistung	5,73	
Buffet	4,36	*
Atmosphäre	5,26	
Standort/Erreichbarkeit	5,35	
Filmangebot	5,52	
Informationsangebot	4,53	
Reservierungsmöglichkeiten	5,04	*
Erscheinungsbild/Image	4,60	*
Film	5,78	

Tabelle 20: Wichtigkeitsmittelwerte

Es ist erkennbar, dass die vier Items, welche die geringst wichtigen der zwölf Variablen darstellen, mit Ausnahme des Informationsangebotes. Geringe Wichtigkeiten deuten auf bereits stark verfestigte *basic*-Faktoren hin. Sie werden durchgehend erfüllt und nicht mehr beachtet. Wird aber eine Schmerzgrenze der Serviceleistungsqualität überschritten, tritt diese als Standard angenommene Serviceleistung wieder ans Tageslicht und wird mit hohen Wichtigkeitswerten ausgezeichnet. Beim Informationsangebot wird das Ausmaß der Servicequalitätsmissachtung geringer ausgefallen sein.

Eine weitere Analyse verdeutlicht diesen Aspekt. Es existiert für jede Servicevariable ein Minimum der Wichtigkeitsbewertung. Dieses Minimum wird getrennt nach Zufriedenheitskategorie für alle elf Zufriedenheitsvariablen ermittelt. Anschließend werden die Durchschnittswerte für die sechs Zufriedenheitskategorien über alle zwölf Serviceleistungsattribute berechnet. Theoretisch liegt das Minimum höher, gewinnt also an Wichtigkeit, je schlechter die Zufriedenheitsbewertungen für die Variablen ausfallen. In Tabelle 21 ist eine

Tendenz zu erkennen, dass bei sinkender Zufriedenheit, die geringst gewählte Wichtigkeit steigt. Dies deutet darauf hin, dass die Spanne der Wichtigkeitsbewertungen mit sinkender Zufriedenheit in die höheren Wichtigkeitsränge steigt und schmaler wird. Diese Erkenntnis bietet letztendlich doch einen kleinen, wenn auch schwachen Hinweis darauf, dass bei sinkenden Zufriedenheitsbewertungen der Wichtigkeitswert steigt.

Zufriedenheitsbewertungskategorie	1	2	3	4	5	6
Wichtigkeitsmittelwerte in den Zufriedenheitsbewertungskategorien	28	31	27	23	25	23

Tabelle 21: Ranking der Wichtigkeitsmittelwerte

Abschließend werden die sechs Zufriedenheitskategorien mit den jeweiligen Wichtigkeitsmittelwerten der zwölf Items korreliert und in einem Streudiagramm wiedergegeben. Die Korrelation wird rein für Nicht-Lieblingskinobesucher durchgeführt, da diese bis auf wenige Ausnahmen in allen Zufriedenheitskategorien über alle Serviceattribute genug Fälle für die Wichtigkeitsbewertungen aufweisen, um die berechneten Mittelwerte rechtfertigen zu können. Die Korrelation wird zwischen 72 Paaren, elf Variable mal sechs Kategorien, durchgeführt und liefert einen Koeffizienten von 0,123 mit einer Irrtumswahrscheinlichkeit von 30,4 %. Es sind keine generalisierbaren Tendenzen zwischen den Serviceleistungsattributen nachweisbar.

Diagramm 37 zeigt die verwendeten Daten in einem Scatterplot. Je geringer die Zufriedenheit, desto eher zeigt sich in manchen Serviceleistungsattributen, wie zuvor festgestellt, die Tendenz höhere Wichtigkeiten aufgrund extrem negativen Empfindens zu nennen. Dieser Aspekt wird durch die kategorial steigende horizontale Streuung innerhalb der jeweiligen Zufriedenheitskategorien dargestellt.

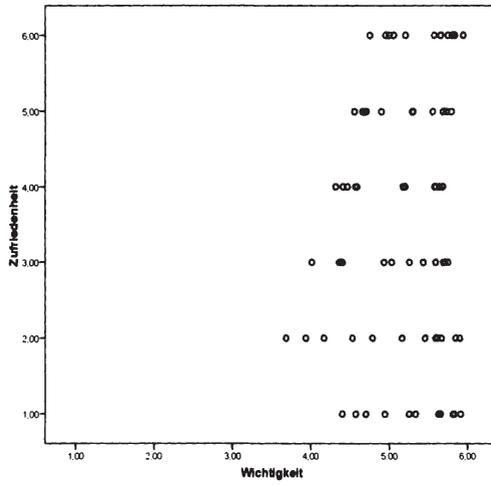


Diagramm 37: Zufriedenheit und Wichtigkeit

4.3. Neue methodische Herangehensweisen

4.3.1. Paarvergleichsmodelle

Ziele:

Die Veränderung der Wichtigkeitsbewertungen wird statisch und über die Zeit hinweg untersucht.

Variablen:

Wie wichtig sind Ihnen folgende Merkmale eines Kinos? (gemessen in den Monaten September 2006 und August 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Wenn Sie an Ihre Kinobesuche denke, seit Sie an der CINEMON-Studie teilnehmen, waren Sie dann ...?“ (gemessen im Monat März 2007)

Items (#2): 1) „... in den meisten Fällen im selben Kino?“; oder 2) „... eher in verschiedenen Kinos?“;

Skalierung: kategorial (1 oder 2);

Programme:

R (Package: prefmod);

Modell:

Hatzinger und Mazanec (2007) verwenden ein *extended Bradley-Terry-Modell* (EBTM) für die Analyse von *conjoint*-Daten. Das EBTM ist ein multinomiales log-lineares Modell aus der Familie der *Generalized Linear Models* und wird über die Kontingenztabelle mittels *maximum likelihood* geschätzt. Sie untersuchen die empfundene Nutzenveränderung von Reisenden durch die Hinzunahme eines Transportmittels, Flug, Zug, oder Bus, zu einem pauschalen Städtebesuchspaket, mit den Attributen Destination, Transportmittel, Art der Übernachtung, Reisedauer und Preis. Die Respondenten bewerten unterschiedliche künstliche Reisepakete auf einer 10-Punkt-Likert-Skala nach deren Buchungswahrscheinlichkeit. Da die metrische Eigenschaft derartiger Ratingskalen angezweifelt wird, verwenden sie Präferenzbeziehungsdaten zwischen Paaren an Reisealternativen, welche mit ‚preferred‘, ‚no preference‘, ‚not preferred‘ ausgewiesen werden. Durch die Recodierung der Daten in Paarvergleiche wird die Anzahl an Ratingskalenpunkten ignoriert. Hatzinger und

Mazanec (2007) vergleichen das Modell mit einem erweiterten Modell, welches die Rangunterschiede in den Präferenzdaten mit einbezieht, finden jedoch keine nennenswerten Unterschiede in den Ergebnissen. Somit scheint die Vernachlässigung dieser Unterschiede auch für den vorliegenden Kinodatenatz nicht von fundamentaler Relevanz zu sein und dient hier als Rechtfertigung der Datenrecodierung. Den Vorteil der log-linearen Formulierung sehen sie in der Möglichkeit, Subjekt und Objektparameter mitschätzen zu können. Subjekte sind in diesem Falle die Respondenten und Objekte die Serviceleistungen. Subjektkovariate können den Einfluss unterschiedlicher Respondententypen auffangen. Ebenso können Interaktionen zwischen Subjekt- und Objekteigenschaften in das Modell eingebaut werden.

Verwendet wird ein R-Paket namens *prefmod* zur Durchführung des log-linearen Bradley-Terry-Modells (Dittrich und Hatzinger, 2008), um die Rankings über Items zu schätzen. Zwischen den Bewertungen verschiedener Items wird Unabhängigkeit angenommen. Im später hier gerechneten longitudinalen Modell wird letzte Annahme aufgelockert. Berechnungen können auch mit durchgehender Bevorzugung eines Items durchgeführt werden. Das heißt, dass beim Vergleich zweier Items immer eines bevorzugt werden muss. Aber auch Modelle mit indifferenten Bewertungen sind schätzbar. Dies bedeutet, dass beim Vergleich zweier Items sowohl die Bevorzugung eines der Items, als auch die Gleichstellung der Wichtigkeit zweier Items möglich ist. Dabei kann zusätzlich ein Parameter berechnet werden, ob eher die Tendenz besteht eine Bevorzugung abzugeben oder sich für eine Gleichstellung der Wichtigkeiten zu entscheiden. Auch Subjektkovariate können eingeführt werden, welche die Rankings durch Personeneigenschaften beeinflussen. Diese Modelle unterstellen, dass nicht alle Personen des untersuchten Datensatzes die gleiche Rangordnung der Items aufweisen und es Unterschiede zwischen den durch die Subjektkovariate eingeführten Gruppen gibt. Ebenso können Interaktionen nicht nur zwischen Items und Subjekten bestehen, sondern beispielsweise auch zwischen Subjektkovariaten und der zuvor genannten Indifferenzhaltung. Weiters können Objektkovariate eingeführt werden, wobei hier gemeinsame Eigenschaften der Items die Bevorzugung oder Gleichstellung einzelner Items beeinflussen. Der Modellfit der vorgestellten Möglichkeiten wird über die Residuendevianz beurteilt, welche einer χ^2 -Quadratverteilung mit zugehörigen Freiheitsgraden folgt.

Der Datensatz besteht vorerst aus 22 Wichtigkeitsbewertungen zu elf Serviceleistungsattributen gemessen zu zwei Zeitpunkten. Da keine direkten Vergleiche abgefragt wurden, nämlich welches der jeweils gewählten

Serviceitems wichtiger ist als ein anderes, müssen diese Vergleiche erst errechnet werden. Dazu werden jeweils immer zwei Variablen in Betracht gezogen. Somit ergeben sich laut Formel 13 sechs verschiedene Kombinationen.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{4!}{(4-2)!2!} = 6$$

Formel 13: Paarvergleichskombinationsbewertungen

Hat Variable eins zum Zeitpunkt eins (November 2006) einen höheren Wichtigkeitswert auf der 6-Punkt-Likert-Skala erhalten als Variable zwei, so erhält die Vergleichsvariable zwischen Variable eins und zwei den Wert 1. Hat Variable eins zum Zeitpunkt eins (November 2006) einen niedrigeren Wichtigkeitswert erhalten als Variable zwei, so erhält die selbige Vergleichsvariable zwischen Variable eins und zwei den Wert -1. Haben die Variablen eins und zwei zum Zeitpunkt eins, November 2006, den gleichen Wichtigkeitswert erhalten, so erhält die selbige Vergleichsvariable zwischen Variable eins und zwei den Wert 0. Dieser Vergleich wird für vier ausgewählte Variablen durchgeführt, welche einer vorangehenden Untersuchung der Wichtigkeitsmittelwerte und Streuung um diesen entspringen. Aus Visualisierungsgründen wird nämlich eine unterschiedliche Variablenvielfalt in das Modell eingebracht. Jede der vier Variablen wird mit den drei anderen verglichen und daraus werden die jeweiligen Vergleichsvariablen gezogen. Im Anschluss wird die gleiche Prozedur für Zeitpunkt zwei (September 2007) vorgenommen.

4.3.1.1. Statische Paarvergleichsmodelle anhand der Kinodaten

Aus den Vergleichen der Wichtigkeitsbewertungen im November 2006 werden zwischen den Variablen die Wichtigkeitsvergleichsbewertungen erzeugt. Die Wichtigkeitsbewertung wird letztendlich in Prozentwerten für die vier Variablen ausgegeben. Im Modell werden die Variablen Buffet, Informationsangebot, Reservierungsmöglichkeiten und Erscheinungsbild/Image betrachtet. Die vorgestellte Paarvergleichsmethode stellt hier einen Vorschlag einer weiteren Möglichkeit dar, Wichtigkeitsbewertungen einzelner Servicekriterien zu ermitteln. Werden die Wichtigkeiten im direkten Vergleich abgefragt, stellt dies eine Erleichterung der Antwortgebung für den Respondenten dar. Die Wichtigkeitswerte werden mit Hilfe des logarithmischen Übergangs und deren zugehörigen *log odds* zwischen ‚Bevorzugung Variable eins‘ und ‚Bevorzugung Variable zwei‘ errechnet. Weitere externe Variable, wie objektspezifische oder subjektspezifische, können, wie auch in erwähnter Studie untersucht, einen Einfluss auf die Struktur der Daten haben und müssen mit Hilfe von Kovariaten

mitgeschätzt werden. Im ersten Modell wird diese Möglichkeit außer Acht gelassen und die gesamtpopulationsspezifischen Wichtigkeitsbewertungen aus der vorhandenen Datenstruktur errechnet. Im ersten Modell ergibt sich eine angenommen χ^2 -verteilte Residuendevianz von 114,5738 mit 8 Freiheitsgraden. Die Wahrscheinlichkeit liegt bei 0 und weist einen schlechten Model-Fit aus. Dies liegt an der Struktur der Daten und möglichen Kovariaten, welche später zwecks Modellverbesserung eingebracht werden. Die Wichtigkeitsmessungen der geschätzten Parameter des Modells in Tabelle 22 weisen signifikante Werte auf. Erscheinungsbild/Image dient als Referenzkategorie mit dem Wert 0. Die Zeile ‚keine Präferenz‘ gibt Auskunft über die Antworttendenz, ob eine Entscheidung getroffen wird, oder ob die beiden Vergleichsattribute als gleich wichtig erachtet werden. Dieser Wert ist nicht signifikant.

Parameter	Estimate	Std.Error	z value	Pr(> z)
Buffet	-0.57863	0.01736	-33.330	< 2e-16 ***
Informationsangebot	-0.23340	0.01652	-14.128	< 2e-16 ***
Reservierungsmöglichkeiten	0.08131	0.01654	4.917	8.79e-07 ***
Erscheinungsbild/Image	NA	NA	NA	NA
„keine Präferenz“	0.01054	0.01671	0.630	0.528

Tabelle 22: Wichtigkeitsparameter (Zeitpunkt eins)

Nach einer Umrechnung der Parameter in Prozentwerte in Tabelle 23 wird klar, dass die Reservierungsmöglichkeiten und das Erscheinungsbild/Image die wichtigsten Attribute der vier darstellen. Das Buffet ist im Vergleich zu den anderen drei Attributen unwichtig und das Informationsangebot liegt im mittleren Bereich. Die Summe der Prozentwerte liegt bei 100. Da die Likert-Skala in Paarvergleiche recodiert wird, geht die absolute Bewertung verloren und es bleiben lediglich die relativen Vergleiche übrig. Somit werden relative Nutzen der Items bestimmt und nicht absolute Bewertungen. Somit können nur Aussagen darüber gemacht werden, welche der ins Modell aufgenommenen Attribute wichtiger und welche unwichtiger, im Vergleich mit den weiteren Attributen im Modell sind, jedoch keine Aussage darüber, ob überhaupt eines der Attribute wichtig bzw. unwichtig ist.

Attribut	Buffet	Informationsangebot	Reservierungs- möglichkeiten	Erscheinungsbild/ Image
Nutzen	10,70	20,11	37,74	32,07

Tabelle 23: Nutzenwerte (Zeitpunkt eins)

Um das Modell zu verbessern wird eine Subjektkovariate eingeführt, nämlich die Selbsteinschätzung der Besucher nach deren Loyalität. Die Respondenten mussten angeben, ob sie ein Lieblingskino besitzen oder nicht. Diese Subjektkovariate beeinflusst möglicherweise das Antwortverhalten der Respondenten. Um dies zu überprüfen wird ein Modell mit einer Subjekt-Objekt-Interaktion geschätzt, welche die Präferenzordnung mitbestimmt. Eine Residuen Devianz von 127,5211 mit 17 Freiheitsgraden zeigt eine Wahrscheinlichkeit von 0 und somit schlechten Modell-Fit. Dies liegt an der Struktur der Daten und möglichen weiteren Subjekt- oder Objektkovariaten, welche hier nicht weiter untersucht werden. Die Wichtigkeitsmessungen der geschätzten Parameter des Modells weisen signifikante Werte in Tabelle 24 auf. Erscheinungsbild/Image mit dem Wert null dient als Referenzwert. Die Zeilen mit Doppelpunkt beschreiben die Interaktionsparameter mit der Subjektkovariaten. Sie sind alle signifikant. Zwei der Interaktionseffekte sind negativ, einer positiv. Negative Werte bedeuten, dass die Wichtigkeit, verglichen mit den anderen Attributen, durch den Einfluss der Loyalitätseigenschaft sinkt. Die relative Wichtigkeit des Informationsangebotes steigt aufgrund der Loyalitätseigenschaft, verglichen mit den anderen Attributen.

Parameter	Estimate	Std.Error	z value	Pr(> z)
Buffet	-0.54288	0.02098	-25.873	< 2e-16 ***
Informationsangebot	-0.26756	0.02024	-13.219	< 2e-16 ***
Reservierungsmöglichkeiten	0.10439	0.02027	5.151	2.59e-07 ***
Erscheinungsbild/Image	NA	NA	NA	NA
„keine Präferenz“	0.01248	0.01672	0.747	0.45535
Buffet:Loyalität	-0.11146	0.03650	-3.054	0.00226 **
Informationsangebot:Loyalität	0.10196	0.03499	2.914	0.00357 **
Reservierungsmöglichkeiten:Loyalität	-0.06905	0.03511	-1.967	0.04921 *
Erscheinungsbild/Image:Loyalität	NA	NA	NA	NA

Tabelle 24: Wichtigkeitsparameter - Subjektkovariate (Zeitpunkt eins)

Die Berechnung der Nutzenwerte in Tabelle 25, zeigt ein ähnliches Bild, verglichen mit dem Modell ohne Einfluss der Subjektkovariate.

Attribut	Buffet	Informationsangebot	Reservierungsmöglichkeiten	Erscheinungsbild/Image
Nutzen	10,08	18,56	39,05	31,69

Tabelle 25: Nutzenwerte - Subjektkovariate (Zeitpunkt eins)

Der Vergleich zwischen den beiden Modellen ergibt eine Devianz von 12,9473 (127,5211-114,5738) bei 9 (17-8) Freiheitsgraden. Laut χ^2 -Verteilung ergibt sich daraus eine Irrtumswahrscheinlichkeit von 16,50 % und das komplexere Modell wird zugunsten des einfachen Modells verworfen. Die Subjektkovariate wird deshalb aus dem Modell entfernt, obwohl sie theoretisch in Interaktion mit weiteren Kovariaten in der Präferenzordnung eine Rolle spielen kann.

Im nächsten Modell wird die Interaktion zwischen der *undecided*-Kategorie, sprich der Kategorie ohne Bevorzugung eines Attributes und der Subjektkovariaten Loyalität eingebracht. Die Wichtigkeitsparameter finden sich in Tabelle 26, die Nutzenwerte in Tabelle 27, wobei diese wiederum ein ähnliches Bild zeigen.

Parameter	Estimate	Std.Error	z value	Pr(> z)
Buffet	-0.54809	0.02120	-25.848	< 2e-16 ***
Informationsangebot	-0.27013	0.02037	-13.262	< 2e-16 ***
Reservierungsmöglichkeiten	0.10539	0.02037	5.174	2.29e-07 ***
Erscheinungsbild/Image	NA	NA	NA	NA
„keine Präferenz“	0.04067	0.02035	1.999	0.04559 *
„keine Präferenz“:Loyalität	-0.08625	0.03573	-2.414	0.01579 *
Buffet:Loyalität	-0.09385	0.03702	-2.535	0.01125 **
Informationsangebot:Loyalität	0.10766	0.03490	3.085	0.00204 **
Reservierungsmöglichkeiten:Loyalität	-0.07072	0.03495	-2.024	0.04301 *
Erscheinungsbild/Image:Loyalität	NA	NA	NA	NA

Tabelle 26: Wichtigkeitsparameter - Subjektkovariate - Interaktion undecided-Kategorie (Zeitpunkt eins)

Das Modell mit der zusätzlichen Interaktion zwischen der Subjektkovariaten und der *undecided*-Kategorie, zeigt eine Residuendevianz von 121,6775 mit 16 Freiheitsgraden, einer folgenden Wahrscheinlichkeit von 0 und somit schlechten Modell-Fit. Dies liegt an der Struktur der Daten und möglichen weiteren Subjekt- oder Objektkovariaten, welche hier nicht mehr weiter untersucht werden.

Attribut	Buffet	Informationsangebot	Reservierungs- möglichkeiten	Erscheinungsbild/ Image
Nutzen	10,06	18,49	39,18	31,73

Tabelle 27: Nutzenwerte – Subjektkovariate - Interaktion undecided-Kategorie (Zeitpunkt eins)

Es ist jedoch erkennbar, dass die Kategorie der Nichtbevorzugung eines Attributes, als auch die Interaktion zwischen dieser Kategorie und der Subjektkovariate, signifikant sind. Deshalb wird mit dem Ausgangsmodell verglichen. Es ergibt sich eine Devianz von 5,8411 (127,5211-121,68) bei einem (17-16) Freiheitsgrad. Laut χ^2 -Verteilung folgt eine Irrtumswahrscheinlichkeit von 1,57 % und das schlankere Modell wird zugunsten des komplexeren Modells verworfen.

4.3.1.2. Longitudinale Paarvergleichsmodelle - Kinodaten

Im letzten Modell wird zusätzlich die Veränderung zwischen den Wichtigkeitsbewertungen von November 2006 bis September 2007 gemessen. Dazu werden die Antworten ein und derselben Person zu zwei Zeitpunkten verglichen. Es wird davon ausgegangen, dass zwischen der Bewertung im November 2006 und September 2007 ein Zusammenhang besteht. Weiters wird davon ausgegangen, dass die Bewertungen einer Person zeitlich voneinander nicht unabhängig sind. Wird also ein Attribut mit zwei unterschiedlichen Attributen in zwei Bewertungen verglichen, wird angenommen, dass die Bewertung des Attributes einer gewissen Standhaftigkeit entspricht. Die Zusammenhänge zwischen den Bewertungsvergleichen bestehen innerhalb der beiden Zeitpunkte, als auch über die Zeit hinweg und werden in das Modell aufgenommen.

Dittrich et al. (2008) schlagen mehrere Möglichkeiten für diesen Zusammenhang vor. Es muss davon ausgegangen werden, dass eine Person, welche Objekte zu einem Zeitpunkt bewertet, diese zum nächsten Zeitpunkt auf die gleiche Art und Weise bewertet. Die Parameter der Zeitabhängigkeit führen ebenfalls zu *log-odds* und *log-odds ratios*. Wie im multivariaten Paarvergleichsmodell wird den Nutzenparametern der Objekte eine zugrundeliegende latente Skala zu jedem Zeitpunkt unterstellt. Abhängigkeiten werden über die Evaluierung gleicher Objektpaare des gleichen Respondenten über unterschiedliche Zeitpunkte miteinbezogen. Der Einfachheit halber wird von der Unabhängigkeit unterschiedlicher Objektpaare ausgegangen, was jedoch über nicht-zeitliche Abhängigkeiten berücksichtigt werden könnte. Es wird eine Markov-Struktur angenommen, welche besagt, dass lediglich die vorangehende Entscheidung einen Einfluss auf den aktuellen Vergleich zweier Objekte hat. Für Abhängigkeitsstrukturen über mehr als zwei Zeitpunkte, eröffnen sich weitere Möglichkeiten als die zuvor genannte, nämlich eine autoregressive Struktur erster Ordnung oder Strukturen höherer Ordnungsabhängigkeiten. Ein Vorteil wird in der Vergleichbarkeit der unterschiedlichen Abhängigkeitstypen über den Devianzvergleich der genesteten Modelle gesehen. Nicht-zeitliche Abhängigkeiten zwischen Paarvergleichsantworten innerhalb eines Zeitpunktes

könnten ebenfalls eingeführt werden, was bedeutet, dass ein Paarvergleich einen anderen beeinflusst, da ein Item in beiden Vergleichen auftaucht.

Um im Kinobeispiel zu ermitteln, ob die Interaktionen über die Zeit weggelassen werden sollen, werden die beiden Modelle in Tabelle 28 verglichen. Die Interaktionen über die Objekte zu den jeweiligen Zeitpunkten sind in beiden Modellen enthalten.

	Devianz	Freiheitsgrade
Modell ohne Interaktionen über die Zeit	7865,927	11240
Modell mit Interaktionen über die Zeit	6163,784	11234
Unterschied	1702,143	6

Tabelle 28: Modellvergleich der zeitabhängigen Modelle

Wird die zeitliche Interaktion entfernt, führt dies zu einer signifikanten Devianzzunahme mit Signifikanzwert 0. Somit wird das simple Modell zu Gunsten des komplexen Modells verworfen. In Tabelle 29 finden sich die geschätzten Parameter. Diese sind alle signifikant. Auch die Interaktionsparameter sind signifikant, was wiederum auf die Bevorzugung des komplexeren Modells hindeutet. Das *odds ratio* des *theta*-Parameters ‚Buffet-Informationsangebot-zeitliche Interaktion‘ berechnet sich aus $\exp(4 \cdot 0,59477) = 10,79497$. Dieser Wert deutet darauf hin, dass die Wahrscheinlichkeit zu beiden Zeitpunkten die gleiche Antwort zu geben höher ist, verglichen mit der Wahrscheinlichkeit eines Wechsels der Präferenzordnung zweier Attribute. Die Berechnung der weiteren *odds ratios* erfolgt analog. Da sämtliche Werte positiv sind, ist eine gleichbleibende Bewertung am wahrscheinlichsten. Die *odds ratios* für einen Wechsel liegen zwischen $\exp(4 \cdot 0,59477) = 10,79497$ und $\exp(4 \cdot 0,14685) = 1,799304$. Die Wechselwahrscheinlichkeit des Vergleiches ‚Informationsangebot-Erscheinungsbild/Image-zeitliche Interaktion‘ liegt niedriger, verglichen mit der Wechselwahrscheinlichkeit von ‚Buffet-Informationsangebot-zeitliche Interaktion‘. In Tabelle 30 und Diagramm 38 finden sich die Nutzenwerte des favorisierten Modells für beide Zeitpunkte.

Parameter	Estimate	Std.Error	z value	Pr(> z)
Buffet-Zeitpunkt 1	-0.24430	0.01448	-16.873	< 2e-16 ***
Informationsangebot-Zeitpunkt 1	-0.12087	0.01338	-9.036	< 2e-16 ***
Reservierungsmöglichkeiten-Zeitpunkt 1	0.05458	0.01372	3.978	6.96e-05 ***
Erscheinungsbild/Image-Zeitpunkt 1	0	NA	NA	NA
Buffet-Zeitpunkt 2	-0.16774	0.01518	-11.051	< 2e-16 ***
Informationsangebot-Zeitpunkt 2	-0.06824	0.01407	-4.849	1.24e-06 ***
Reservierungsmöglichkeiten-Zeitpunkt 2	-0.03937	0.01437	-2.739	0.00617 **
Erscheinungsbild/Image-Zeitpunkt 2	0	NA	NA	NA
Buffet-Informationsangebot-zeitliche Interaktion	0.59477	0.03462	17.180	< 2e-16 ***
Buffet-Reservierungsmöglichkeiten-zeitliche Interaktion	0.50267	0.03633	13.837	< 2e-16 ***
Buffet-Erscheinungsbild/Image-zeitliche Interaktion	0.39366	0.03310	11.894	< 2e-16 ***
Informationsangebot-Reservierungsmögl.-zeitliche Interaktion	0.46135	0.03558	12.967	< 2e-16 ***
Informationsangebot-Erscheinungsbild/Image-zeitliche Interaktion	0.14685	0.03191	4.603	4.17e-06 ***
Reservierungsmögl.-Erscheinungsbild/Image-zeitliche Interaktion	0.47755	0.03308	14.436	< 2e-16 ***

Tabelle 29: Geschätzte Parameter des favorisierten zeitabhängigen Modells

Zeitpunkt	Buffet	Informationsangebot	Reservierungsmöglichkeiten	Erscheinungsbild/Image
1	19,87 %	24,10 %	28,65 %	27,38 %
2	20,61 %	24,90 %	27,19 %	27,30 %

Tabelle 30: Nutzenwerte (beide Zeitpunkte)

Es ist erkennbar, dass sich die Wichtigkeitswerte der vier Variablen über die Zeit verändern. Die Reservierungsmöglichkeit wird zum Zeitpunkt eins als wichtigstes Attribut der vier betrachtet, zum Zeitpunkt zwei fällt dies jedoch an die zweite Stelle zurück und Image ist das wichtigste Attribut. Buffet ist zu beiden Zeitpunkten das unwichtigste Attribut. Das Informationsangebot liegt im

Ranking zu beiden Zeitpunkten an vorletzter Stelle. Es darf jedoch nicht interpretiert werden, dass Buffet an Wichtigkeit gewinnt, da der Punkt zum Zeitpunkt zwei höher liegt. Schließlich geben diese Werte keine Auskunft über die absolute Wichtigkeit. Es kann lediglich die Schlussfolgerung gezogen werden, dass die Wichtigkeitstrennung zum Zeitpunkt eins größer war, verglichen mit Zeitpunkt zwei. Die Wichtigkeiten der Attribute werden sich somit über die Zeit hinweg ähnlicher.

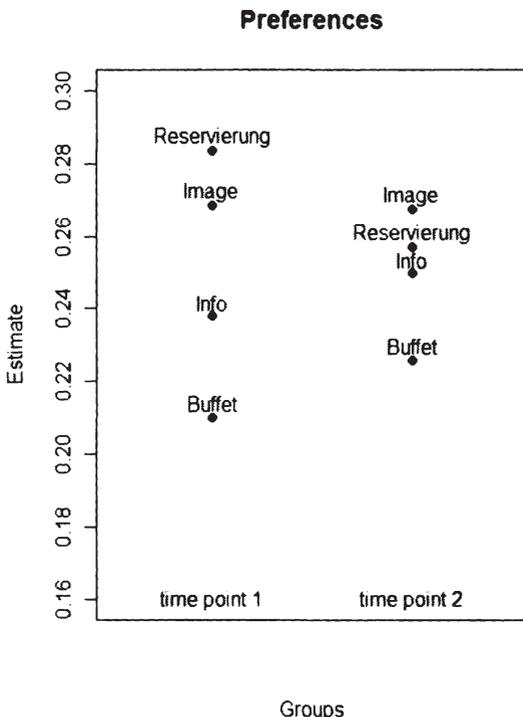


Diagramm 38: Nutzenwerte (beide Zeitpunkte)

Um graphisch darzustellen wie weit diese von einer angenommenen gleichbleibenden Wichtigkeitsbewertung abweichen, wird in Diagramm 39 eine Gerade mit einer 45° Steigung durch die geplotteten Wichtigkeitsbewertungen gezogen. Auf der x-Achse sind die Wichtigkeitswerte der vier Variablen für Welle zwei im November 2006 aufgetragen. Auf der y-Achse bestimmen die zugehörigen Wichtigkeitswerte für Welle zwölf im September 2007 die Höhe der Lage der Punkte.

Es zeigt sich, dass nur Erscheinungsbild/Image (E) über die Zeit hinweg die relativ im Vergleich mit den anderen drei Variablen gesehene Wichtigkeit, nicht ändert. Die Variablen Erscheinungsbild/Image und Reservierungsmöglichkeiten (R) weisen weiters die höchsten Wichtigkeitswerte auf. Der Wichtigkeitswert der Variablen Reservierungsmöglichkeiten sinkt im Intervall der zehn dazwischen liegenden Monate verglichen mit den relativen Wichtigkeiten der weiteren Variablen, wo hingegen das Informationsangebot (I), relativ zu den anderen drei Variablen, an Wichtigkeit gewinnt. Die Wichtigkeit der Variablen Buffet (B) weist relativ gesehen ebenfalls einen leichten steigenden Trend im Vergleich zu den weiteren Attributen auf. Es muss hier jedoch nochmals beachtet werden, dass die sich die Wichtigkeit über alle drei Variablen immer auf eins aufsummiert. Somit dürfen keine Aussagen darüber gemacht werden, ob Variablen an Wichtigkeit über die Zeit hinweg gewinnen oder verlieren. Aussagen dürfen lediglich in Relation zu den weiteren im Modell enthaltenen Variablen gemacht werden.

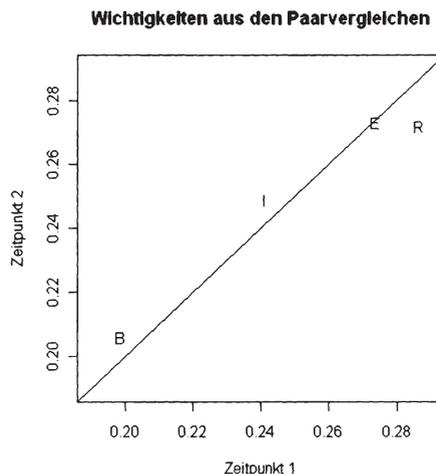


Diagramm 39: Paarvergleichswichtigkeiten im zeitabhängigen Vergleich

Eine genauere Beschreibung des Modells inklusive R-Code und Daten finden sich einem Werk von Abu Ghannam et al. (2009). Darin ist auch eine Erweiterung des zuletzt bevorzugten zeitabhängigen Modells mit der Subjektkovariaten Loyalität enthalten, wobei hier interessante Unterschiede zwischen den Loyalitätsgruppen aufgedeckt werden konnten.

4.3.2. *Item Response Theory: vom Rasch model zum linear logistic model with relaxed assumptions*

Ziele:

Die innewohnenden Probleme der Datenstruktur werden aufgedeckt und Verbesserungsvorschläge für zukünftige Befragungen vorgenommen. Die Schlussfolgerungen dienen als Hinweis dafür, die darauffolgenden Analysen der Arbeit mit Vorsicht zu interpretieren. Auch eine longitudinale Verwendung der *item response theory* wird vorgestellt.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den Monaten Oktober 2006 bis März 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Wie wichtig sind Ihnen folgende Merkmale eines Kinos? (gemessen in den Monaten September 2006 und August 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr wichtig) bis 1 (nicht wichtig);

Programme:

R (Package: eRm); Rumm2020;

Modell:

Rasch-Modelle zählen zu den probabilistischen Testtheorien, den *latent trait*-Modellen. Bühner (2006) bietet eine einfache Einführung in die Eigenheiten der Rasch-Methodologie. Im Unterschied zu den *item response theory* (IRT) basierten Modellen, geht die *classical test theory* (CTT) davon aus, dass die Werte einer Person zu einer Gesamtsumme aufaddiert werden können und Intervallskalenniveau besitzen, was jedoch nur in den seltensten Fällen angenommen werden kann. Die klassische Testtheorie geht also von sogenannten *tau*-äquivalenten Items aus. Dies sind Items mit gleicher Trennschärfe. Somit ist es gleichgültig, welche Items eine Person z.B. richtig beantwortet. Sie werden ohne weitere Umformung laut der suffizienten

erschöpfenden Statistik aufsummiert. Es gibt keine schwierigeren oder leichteren Items. Davon kann im Normalfall jedoch ebenfalls nicht ausgegangen werden. Bei den Rasch-Modellen werden sowohl die Itemschwierigkeit, als auch die Personenfähigkeit, auf eine gemeinsame latente Skala gebracht, was eine anschließende Aufsummierung der Itemwerte ermöglicht.

Beim klassischen Rasch-Modell wird Eindimensionalität angenommen, warum dieses auch starkes *latent trait*-Modell genannt wird. Die Eindimensionalität setzt lokale Unabhängigkeit voraus, wodurch Korrelationen zwischen Items nur aufgrund einer latenten Dimension entstehen. Lokale Unabhängigkeit bedeutet, dass die Wahrscheinlichkeit einer Antwort auf ein Item nicht von der Antwort eines anderen Items abhängt. Somit ist die bedingte Wahrscheinlichkeit, also die Wahrscheinlichkeit einer Antwort auf ein Item, unter der Bedingung, dass eine gewisse Antwort eines anderen Items gewählt wurde, gleich der Wahrscheinlichkeit ohne dieser Bedingung. Beinhaltet das IRT-Modell Personenparameter nur einer Dimension, ist die lokale Unabhängigkeitsannahme für die Eindimensionalität von Bedeutung, da untersucht wird, ob ein einzelner *latent trait* den beobachteten Fragen zu Grunde liegt. In der CTT werden die *trait levels* durch die Verbindung der Antworten über die Items bewertet. Dies passiert durch Aufsummierung dieser. Der beobachtete Wert ist jedoch die Summe des wahren Wertes und des Fehlerwertes. Die Annahme des wahren Wertes passt nämlich nur zu wenigen Items. Weiters sind die Itemeigenschaften nicht mit dem Verhalten beim Test verknüpft und unabhängige Variablen sind nicht separat messbar. IRT hingegen beschäftigt sich nicht damit, wie die Itemantworten aufaddiert werden können. Vielmehr geht es um die Eigenschaften der Items und wie diese das Verhalten beeinflussen, also welches *trait level* am besten die Antworten beschreibt. So werden *item characteristic curves* (ICCs) an die Daten angepasst, wobei Veränderungen des *trait levels* mit Veränderungen der *item response*-Wahrscheinlichkeiten verknüpft sind. Weiters legt die lokale Unabhängigkeitsannahme fest, dass die Item- und Personenparameter die Zusammenhänge zwischen Items und Personen vollständig beschreiben. Es existieren keine weiteren Beziehungen in den Daten.

Weiters sind Messungen der Itemschwierigkeit und Personenfähigkeit unabhängig voneinander. Dies bedeutet, dass jede Teilstichprobe einer Gesamtstichprobe immer zu den gleichen geschätzten Parametern führt, auch Invarianzeigenschaft oder spezifische Objektivität genannt. Unterschiede zwischen Itemparametern hängen somit nicht von der verwendeten Personenstichprobe ab und Unterschiede zwischen den Personenparametern nicht von der verwendeten Itemstichprobe. In Abgrenzung dazu hängen

Itemantworten in den allgemeinen *item response*-Modellen von mehr als einer latenten Fähigkeit ab, wie z.B. von einem Rate- oder Trennschärfeparameter. Die fundamentalsten Unterschiede zwischen der klassischen und probabilistischen Testtheorie listen Embretson und Reise (2000) in Tabelle 31 auf.

Aus ‚Rule 8‘ ergeben sich Hinweise in Bezug auf longitudinale Messungen. Es resultieren nämlich paradoxe Reliabilitäten. Je niedriger die Korrelation zwischen *pretest* und *posttest*, desto höher ist der Reliabilitätswert der Veränderung. Aufgrund der Subtraktion können sich zwischen dem anfänglichen Status und der Veränderung negative Korrelationen ergeben. Weiters ergeben sich veränderte Bedeutungen, abhängig vom anfänglichen Level, da in der CTT Intervallskalierung laut ‚Rule 6‘ meist nicht gerechtfertigt werden kann. Im IRT Modell, speziell im Rasch Modell, jedoch schon. Veränderung vom IRT *trait level* weg haben auch bei unterschiedlichen anfänglichen Levels eine gleichbleibende Bedeutung. Somit verspricht die Rasch-Modellierung eine genauere longitudinale Veränderungsmessung, verglichen mit Methoden der CTT.

Die klassische Testtheorie widerspricht der quasi perfekten Beantwortung von Fragen. Es dürfte theoretisch nicht passieren, dass eine Person deren Fähigkeit höher liegt als die Schwierigkeit des Items, dieses Item falsch beantwortet. Situationsabhängig tauchen jedoch Abweichungen auf, welche Rasch-Modelle aus der Familie der *logit*-Modelle miteinbeziehen, im Gegensatz zur deterministischen Guttman-Skala, welche hinter der CTT steht. Abweichungen führen dazu, dass der Übergang zweier Antwortmöglichkeiten nicht abrupt passiert, also nicht wie bei einer Sprungfunktion in jenem Punkt genau dort, wo die Itemschwierigkeit der Personenfähigkeit entspricht. Es stellt sich die Frage, ob die Wahrscheinlichkeit zwischen bspw. ‚Ja‘ und ‚Nein‘ überlaufend passiert und ein Kippen der Antwort in der Nähe des Überganges wahrscheinlicher wird. Es wird folgend davon ausgegangen, dass die Wahrscheinlichkeit, z.B. eine Frage richtig zu beantworten, mit der Personenfähigkeit steigt, also nicht deterministisch, sondern probabilistisch bestimmt ist. Der Übergang zwischen den Antwortmöglichkeiten wird Schwelle oder *threshold* genannt. Diese Wahrscheinlichkeitsannahmen werden mit Hilfe der logistischen Funktion gelöst. In den später folgenden Diagrammen wird dies durch sich asymptotisch annähernde Kurven veranschaulicht, auch ICCs genannt. Die ICC-Kurven nehmen einen S-förmigen Verlauf an, welche die Wahrscheinlichkeit einer Antwort auf einer monoton steigenden Funktion abbilden. In der Mitte der Kurve ist die Wahrscheinlichkeitsveränderung einer Antwort bei einer Veränderung der Personenfähigkeit am größten. An den extremen Ausprägungen ist diese

Veränderung am kleinsten. Die Steigung beschreibt die Itemdiskriminierung. Je steiler diese ist, desto schneller kommt es zu einer Antwortwahrscheinlichkeitsveränderung bei einer Änderung der Personenfähigkeit. Es gibt für jede Personenfähigkeit Wahrscheinlichkeitswerte für alle Antwortmöglichkeiten größer null, welche in Summe eins ergeben. In Modellen, in welchen eine Ratewahrscheinlichkeit impliziert wird, liegt die Wahrscheinlichkeit einer Antwort niemals bei null.

	The Old Rules
Rule 1	The standard error of measurement applies to all scores in a particular population.
Rule 2	Longer tests are more reliable than shorter tests.
Rule 3	Comparing test scores across multiple forms is optional when the forms are parallel.
Rule 4	Unbiased estimates of item properties depend on having representative samples.
Rule 5	Test scores obtain meaning by comparing their position in a norm group.
Rule 6	Interval scale properties are achieved by obtaining normal score distributions.
Rule 7	Mixed item formats leads to unbalanced impact on test total scores.
Rule 8	Change scores cannot be meaningfully compared when initial score levels differ.
Rule 9	Factor analysis on binary item produces artifacts rather than factors.
Rule 10	Item stimulus features are unimportant compared to psychometric properties.
	The New Rules
Rule 1	The standard error of measurement differs across scores (or response patterns), but generalizes across populations.
Rule 2	Shorter tests can be more reliable than longer tests.
Rule 3	Comparing test scores across multiple forms is optimal when test difficulty levels vary between persons.
Rule 4	Unbiased estimates of item properties may be obtained from unrepresentative samples.
Rule 5	Test scores have meaning when they are compared for distance from items.
Rule 6	Interval scale properties are achieved by applying justifiable measurement models.
Rule 7	Mixed item formats can yield optimal test scores.
Rule 8	Change scores can be meaningfully compared when initial score levels differ.
Rule 9	Factor analysis on raw item data yields a full information factor analysis.
Rule 10	Item stimulus features can be directly related to psychometric properties.

Tabelle 31: Klassische vs. probabilistische Testtheorie (Embretson und Reise, 2000)

In Modellen, in welchen es mehr als zwei Antwortmöglichkeiten gibt, werden diese Kurven *category characteristic curves* (CCC) genannt. Verglichen mit dem dichotomen Modell, wo der Wahrscheinlichkeitsübergang einer Itemschwierigkeit bei 50 % liegt, liegt die Itemschwierigkeit bei Items mit mehreren

Kategorien im Mittel der geschätzten Schwellen. Es gibt zwei bekannte Modelle mit geordneten Schwellen. Das *rating scale model* (RSM), welches davon ausgeht, dass die Abstände zwischen zwei aufeinander folgenden Schwellen über alle Items gleich sind und somit gleiche Trennschärfe der Items vorausgesetzt wird. Die Items unterscheiden sich dann lediglich in der Itemschwierigkeit. Im Äquidistanzmodell wird zusätzlich der Abstand zwischen allen Schwellen über alle Items als gleich groß angenommen. Im *partial credit model* (PCM) existieren diese Annahmen nicht. Sowohl die Abstände zwischen den Schwellen innerhalb eines Items, als auch über die Items hinweg, dürfen variieren.

Im einfachen dichotomen Modell bestehen lediglich Unterschiede bezüglich der Itemschwierigkeit. Das dichotome Rasch-Modell (1-Parameter-Modell) hat zwei Antwortmöglichkeiten und somit auch nur eine Wahrscheinlichkeit bzw. die zugehörige Gegenwahrscheinlichkeit. Da die Wahrscheinlichkeiten zwischen null und eins liegen, fällt der Übergang zwischen bspw. ‚richtig‘ und ‚falsch‘ auf 50 %. An dieser Schwelle liegt auch gleichzeitig die Itemschwierigkeit auf der latenten Dimension. Die ICC-Kurven können sich nicht überschneiden, da die Steigungen über die Items gleich sind. Dort wo der Übergang zwischen zwei Kategorien stattfindet, also bei einem Wahrscheinlichkeitsverhältnis von 50:50, liegt auch der Schwierigkeitsgrad des Items. Dieser stellt gleichzeitig die einzige Schwelle des Items dar. Da die Standardabweichungen für extrem leichte bzw. schwierige Items höher sind, können diese zur Bestimmung der unterschiedlichen Konfidenzintervalle um den Schwierigkeitsgrad des Items genutzt werden.

Verglichen damit, berücksichtigt das Birnbaum-Modell, *two parameter logistic* Modell, einen zusätzlichen Parameter, nämlich einen Trennschärfeparameter, *item discrimination*, welcher die Steigung der ICC bestimmt. Es erlaubt unterschiedliche *slopes* und sich überschneidende Kurven. Je steiler die Kurve, desto trennschärfer, und vice versa.

Im *three parameter logistic* Modell wird zusätzlich ein Parameter für die Ratewahrscheinlichkeit mitgeschätzt. Dieser verhindert, dass für ein Item, für welches die richtige Antwort einen prozentuellen Schätzwert besitzt, dieser auf eine Wahrscheinlichkeit einer richtigen Antwort von null % sinkt. Die letzten beiden Modelle sind aufgrund der Vielzahl an zu schätzenden Parametern teilweise instabil.

Im Schätzverfahren werden jene Wahrscheinlichkeiten gesucht, welche den höchsten *likelihood*-Wert ausgeben. Gegeben die Itemeigenschaften, wird die

likelihood eines Antwort-patterns für den geschätzten *trait*-Level einer Person maximiert. Da das Antwortverhalten aber meist nicht konsistent ist, kommt es zu Inkonsistenzen, welche den *likelihood*-Wert verringern. Also eine Person mit niedriger Fähigkeit, beantwortet ein Item mit hohem Schwierigkeitsgrad, aber eines mit niedrigem nicht, und vice versa. So kann der *likelihood*-Wert zwischen Personen unterschiedlich sein, auch wenn diese den gleichen aufsummierten Wert aufweisen. Es ist aber ziemlich unwahrscheinlich, dass eine Person mit hoher Personenfähigkeit, Items mit geringem Schwierigkeitsgrad nicht beantwortet, und vice versa. Der *likelihood*-Wert orientiert sich deshalb stark an den am besten diskriminierenden Items, da diese die meiste Information enthalten. Im klassischen Rasch-Modell, wobei von einer erschöpfenden Statistik ausgegangen wird, ist es gleichgültig, welche Items beantwortet wurden. Es gibt darin keinen Diskriminierungskoeffizienten. Im 2-Parameter-Modell, welches einen Parameter für die Diskriminierung beinhaltet, hängen die Personenparameter davon ab, welche Items wie beantwortet wurden. Probleme in der Rasch-Modellierung gibt es z.B. für Personen, welche alle Items richtig oder falsch beantwortet haben. Deren Datenzeilen beinhalten keine Information und somit kommt es zu Schätzverzerrungen. Die Schätzgenauigkeit hängt also von der Verteilung der Itemschwierigkeiten über die Verteilung der Personenfähigkeiten ab. Je näher und enger diese beieinander liegen, desto besser werden die Daten zum Rasch-Modell passen. Dieser Punkt wird später veranschaulicht.

Zuerst werden die Itemparameter, also der Wert der Itemschwierigkeit auf der latenten Dimension, geschätzt und im Anschluss die Personenparameter, also die Personenfähigkeit auf der latenten Dimension. Da auch ordinale Items verrechnet werden können, benötigen diese Modelle verschiedene Annahmen. Das Trennschärfekriterium muss bei diesen Modellen entfernt werden, was schon durch die Skalenbreiten ausgeschlossen wird. Dafür müssen jedoch die Antwortschwellen geordnet sein. Dies bedeutet, dass sich die beantworteten Kategorien sinnvoll zum Anstieg der Personenfähigkeit verhalten müssen. Weiters ist die Schätzfehlervarianz umso geringer, je näher die Itemschwierigkeit bei der Personenfähigkeit liegt, da dann die Fähigkeit der Person exakt gemessen werden kann, quasi am richtigen Punkt getroffen wird. Als Schätzverfahren stehen beispielsweise die *unconditional*, die *marginal* und die *conditional maximum-likelihood* (ML) Methode zur Verfügung (uML, mML, cML). Nachteil der uML ist die gleichzeitige Schätzung der Item- und Personenparameter und bei der mML, dass trotz getrennter Schätzung der Item- und Personenparameter, die Annahme einer Normalverteilung über die Personen angenommen wird. Deshalb wird cML empfohlen, welche die Itemparameter unabhängig von den

Personenparametern schätzt und erst einen objektiven Vergleich zwischen Tests ermöglicht.

Sind die Daten nicht raschkonform, gibt es unterschiedliche Möglichkeiten diesem Problem zu begegnen. Im *mixed* Rasch-Modell (MRM) werden gleichzeitig im Zuge der Schätzung latente Klassen identifiziert, welche sich in den Itemparametern maximal unterscheiden. Daraus ergibt sich ein unterschiedliches Antwortverhalten zwischen den Gruppen, welche folglich auch nicht miteinander verglichen werden können, da die Items in den Gruppen unterschiedliche Einflüsse messen. Sind die Gruppen in sich selbst immer noch nicht raschkonform, stehen unterschiedliche Tests zur Verfügung. Diese dienen dazu, z.B. die Itemformulierung zu verbessern, unpassende Items auszuschließen, oder Probleme einzelner Kategorien innerhalb von Items oder Items selbst zu identifizieren. Ein graphischer Modelltest kann diese Probleme veranschaulichen. Dabei wird die Stichprobe in zwei Teilstichproben geteilt, für beide Stichproben die Itemparameter bestimmt und diese in einem Streudiagramm geplottet. Liegen diese auf der Diagonalen, so sind die Itemparameter ident und raschkonform. Beim *likelihood*-Quotienten-Test wird der *likelihood*-Wert der Daten entweder mit den Teilstichproben (bedingter *likelihood*-Quotienten-Test), oder mit dem saturierten Modell, deren Daten perfekt das Rasch-Modell beschreiben, verglichen. Der Pearson- χ^2 -Test zielt auf die Häufigkeiten der Antwortmuster ab und vergleicht die tatsächlichen mit jenen, welche aus den Annahmen des Raschmodells hervorgehen. Je größer der Wert, desto eher wird Raschkonformität verworfen. Der Wert wird signifikant. Um Modelle untereinander vergleichen zu können, steht beispielsweise der χ^2 -Differenztest für hierarchisch geschachtelte, also genestete Modelle, zur Verfügung. Da komplexere Modelle es leichter vermögen, Daten zu beschreiben, beziehen Informationskriterien die Modellkomplexität mit ein. Die Bedingung der hierarchischen Verschachtelung muss dafür nicht erfüllt sein. Was diese Informationskriterien nicht beurteilen können, ist die Güte der Daten. Sie erlauben keine Aussage darüber, ob das Modell gut ist, sondern lediglich einen Vergleich zwischen zwei Modellen, welches besser ist. Das *Bayes Information Criterion* (BIC) und das *Consistent Akaike Information criterion* (CAIC) werden dem *Akaike Information Criterion* (AIC) vorgezogen. Die Berechnungen sind in Formel 14 aufgeführt.

$$AIC = -2 \log L + 2n_p$$

$$BIC = -2 \log L + (\log N) * 2n_p$$

$$CAIC = -2 \log L + (\log N) * n_p + n_p$$

Formel 14: Informationskriterien

Über die Differenz der ML-Werte, welche die Abweichungen der Daten vom Modell angeben, und die Differenz der Freiheitsgrade, werden zwei Modelle über die χ^2 -*statistic* verglichen. Ist der Wert aus Formel 15 signifikant, wird das komplexere Modell gegenüber dem simpleren Modell verworfen.

$$\chi^2 = -2 \log \text{Likelihood}_{\text{simple}} - (-2 \log \text{Likelihood}_{\text{complex}})$$

Formel 15: Modellvergleich

Werden Items mit *multiple ordered-response categories* verwendet, müssen polytome Modelle herangezogen werden, um eine nicht-lineare Beziehung zwischen der Personenfähigkeit und der Wahrscheinlichkeit der einzelnen Antworten zu ermöglichen. Es kann das *graded-response model* (GRM) von Samejima für geordnet kategoriale Antworten wie Likert Ratingskalen verwendet werden. Die Items müssen nicht die gleiche Anzahl an Antwortkategorien aufweisen. Verglichen werden, wie später dargestellt, jeweils die Paare ‚1‘ vs. ‚2, 3, 4, 5, 6‘; ‚1, 2‘ vs. ‚3, 4, 5, 6‘; ‚1, 2, 3‘ vs. ‚4, 5, 6‘; etc. Das *modified* GRM (M-GRM), ein eingeschränktes GRM, erlaubt *item-slope* Parametern Variation über die Items und somit wird für jedes Item ein Set an Kategorienschwellen geschätzt. Im GRM gibt es nur eines. Eingeschränkt ist das Modell deshalb, weil die Kategoriengrenzen über die Items gleich weit voneinander entfernt sind, im GRM können diese jedoch über die Items variieren. Das M-GRM benötigt somit weniger Parameter. Das Problem unterschiedlicher Itemskalen kann aber auch anders gelöst werden. Dazu werden die Items mit gleicher Skala in Blocks zusammengefasst und die Parameter blockweise berechnet. Die Blocks sind untereinander nicht vergleichbar. Das PCM von Masters (1982) kann über die Beantwortung eines Item über mehrere Stufen erklärt werden. Die Respondenten erhalten *partial credit* für die Erfüllung mehrerer Schritte im Lösungsprozess. Für jede Personenfähigkeit kann die erwartete Antwort bestimmt werden, nämlich jene, wo die Antwortwahrscheinlichkeit am höchsten liegt. Die Antworten müssen nicht notwendigerweise geordnet sein. Dies wird *reversal* genannt und kommt vor allem dann vor, wenn in einer Kategorie eines Items kaum Antworten auftauchen. Somit kann es vorkommen, dass diese Kategorie niemals die am wahrscheinlichsten auf der Personenfähigkeitsdimension ist. Deshalb wird

empfohlen eine Stichprobe zu finden, welche möglichst heterogen ist, um die Verwendung aller Kategorien zu garantieren. Gelingt dies nicht, so können keine guten Schwellenwerte geschätzt werden. Gründe dafür kann ein falsches Einschätzen der latenten Dimension sein. Lösungen dafür stellt die Zusammenlegung von Kategorien, oder eine erneute Datensammlung mit weniger Kategorien im Problemitem dar. So wie das PCM, diskriminieren auch im RSM alle Items gleich und der reine Skalenwert ist eine suffiziente Statistik um den *trait*-Level der Personen zu messen. Im Vergleich zum RSM, geht das PCM jedoch nicht davon aus, dass die Stufen, also die Kategorien eines Items, gleich schwer über die Items sind. Somit können im PCM die Kategorien über die Items variieren. *Divide-by-total* oder *direct* IRT Modelle können Itemantworten beschreiben, welche über das *trait*-Kontinuum nicht geordnet sind (Bock, 1972). Im Folgenden werden die drei genannten Modelle genauer beschrieben.

Andrich (1978) geht von den dichotomen Modellen aus, nämlich dem Birnbaum Modell von Lord und Birnbaum und dem Rasch Modell von Rasch und Andersen, und formuliert ein Modell für *polytomous response models for unidimensional traits*, wobei z.B. die Einstellung eines Subjektes und der affektive Wert eines Items Gegenstand der Untersuchung sein können. Zuerst wird das Modell von dichotom auf trichotom erweitert und dann eine allgemeine Notation für eine beliebige Anzahl an Kategorien vorgestellt, welche jedoch über die Items gleich sein muss. Dieses Modell wird als RSM bezeichnet. Im Artikel wird von geordneten Antwortkategorien ausgegangen. Das heißt, pro Item gibt es eine Antwortmöglichkeit, sie schließen sich gegenseitig aus und den Antwortmöglichkeiten wird eine geordnete Struktur zugrunde gelegt. Um den affektiven Wert eines Items zu bestimmen, benötigen die *thresholds* so etwas wie Gewichtungen. Fließen diese nicht in das Modell ein, wäre eine Zustimmung zu einem Item mit einem moderaten affektiven Wert, gleichbedeutend mit der neutralen Haltung zu einem Item mit einem hohen affektiven Wert. Um die quasi sequentiellen Antworten eines Items zu teilen, werden die Wahrscheinlichkeiten der Antwortmöglichkeiten als Teilprozesse angeschrieben. Da diese Teilprozesse jedoch einander ausschließende Kriterien eines Items sind, werden diese wieder zusammengeführt. Der Proband wird mit einer konfliktären Situation befasst, für deren Lösungen die Bedingungen bekannt sind. Entscheidet sich der Proband, hat er den Konflikt gelöst. Zusammengefasst wird das ursprüngliche Verständnis, wo es um das Erreichen eines gewissen Levels geht, für eine Entscheidungssituation umformuliert und mathematisch notiert. In einem weiteren Schritt wird spezifiziert, dass jeder *threshold* gleich diskriminiert, gleichgültig zu welchem Item dieser gehört. So kann der Level einer Person auf

der latenten Dimension, über die suffiziente Statistik, also die Summe der überwundenen *thresholds*, bestimmt werden. Im dichotomen Fall wäre dies die Anzahl an erfüllten Items. Gibt es nur ein Item, zählt wiederum die Anzahl an überwundenen *thresholds*. Andrich erweitert dieses Modell um die Einschränkung equidistanter *thresholds*, dies bedeutet gleiche Distanzen zwischen den Schwellen.

Masters (1982) erweitert das RSM von Andrich, wobei die Ordnung und die Anzahl an Kategorien über die Items variieren können. Dies setzt die Generalisierung der Skalen über die Stichproben voraus, also eine suffiziente Statistik. Es wird wiederum vom klassischen Rasch-Modell ausgegangen, welches die Berechnung der Itemparameter unabhängig von der Berechnung der Personenparameter ermöglicht. Er beschreibt das Modell für geordnete Kategorien, wiederholte Versuche bei welchen es eine fixe Anzahl an Versuchen gibt und die Erfolgsrate gezählt wird, Zähldaten bei welchen es keine Obergrenze gibt und Ratingskalen. Letztere Modellklasse trifft auf die vorliegende Kinostudie zu. *Partial credit* bedeutet die Möglichkeit eines Teilerfolgs verglichen mit dichotomen Skalen, wo es nur die beiden Ausprägungen Erfolg vs. Misserfolg gibt. Somit wird das Modell auch PCM genannt. Zuerst wird Samejima's GRM beschrieben, in welchem die Itemparameter nicht unabhängig von den Personenparametern berechnet werden können und deshalb keine spezifische Objektivität gewährleistet wird. Es wird die Sichtweise eingenommen, dass zuerst die unteren Levels erfüllt werden müssen, also die unteren *thresholds* überschritten werden müssen, um eine Antwort in einer höheren Kategorie geben zu können. Somit werden auf einer Skala mit vier Antwortmöglichkeiten, null, eins, zwei oder drei, nicht drei Schwierigkeiten berechnet (Schwierigkeit Antwort eins zu geben, Schwierigkeit Antwort zwei zu geben, und Schwierigkeit Antwort drei zu geben), sondern die Skala wird als stufenförmiger Prozess betrachtet. Somit wird beispielsweise die Wahrscheinlichkeit einer Person, welche mit zwei geantwortet hat, daraus berechnet, wie wahrscheinlich es ist, drei zu antworten. Dies hat also nichts mit der Antwortwahrscheinlichkeit zwei zu tun. Diese hängt nämlich nur von der Fähigkeit der Person ab, wie schwer es für sie ist, die ersten beiden Schritte des Items zu gehen. Eben erläuterte Punkte werden später in Diagramm 42 visualisiert. Letztendlich hängt die Wahrscheinlichkeit natürlich auch von der Schwierigkeit der anderen Items ab, jedoch nicht von der Person selbst, also wiederum eine suffiziente Statistik. Der Unterschied liegt letztendlich darin, dass in Samejima's GRM der *threshold* für den Sprung zwischen den Kategorien bei der 50 %-igen Wahrscheinlichkeit der Vorkategorie liegt, im PCM liegt dieser beim Schnittpunkt der beiden Kategorien. Da in Samejima's GRM eine Ordnung

angenommen wird, welche nicht immer vorausgesetzt werden kann, da zum Beispiel die Möglichkeit besteht Kategorien zu überspringen, muss mit kumulativen Wahrscheinlichkeiten gerechnet werden, was die suffiziente Statistik verbietet. Im PCM errechnet sich die Fähigkeit einer Person aus der Anzahl an erreichten *threshold*-Sprüngen. Die Itemschwierigkeit errechnet sich aus der Anzahl an Personen, welche einen gewissen *threshold*-Sprung vollzogen haben.

Bock (1972) beschreibt ein *latent trait model*, nämlich das *nominal response model*, für Items mit zwei oder mehr nominalen Kategorien. Wird beispielsweise eine Multiple-choice Frage falsch beantwortet, kann höchstwahrscheinlich nicht davon ausgegangen werden, dass die Antworten einheitlich über die falschen Antwortmöglichkeiten streuen. Die Information, welche den falschen Antworten zugrunde liegt, wird zur Bestimmung der Personenfähigkeit herangezogen. Bock (1972) stellt die Methode der cML als auch der uML vor. Bei der Berechnung der Itemschwierigkeiten wird bei uML eine zugrunde liegende Population an Personenfähigkeiten unterstellt. Die Stichprobe wird als zufällig von der Population gezogen betrachtet. Die Itemparameter werden geschätzt und im Anschluss daran werden diese verwendet, um die Personenfähigkeiten für die zugrunde liegende Population zu schätzen. Im Gegensatz dazu werden bei cML die Personenfähigkeiten und die Itemschwierigkeiten simultan geschätzt. Somit sind diese nicht an eine unterstellte Population gebunden. Die cML weicht somit die lokale Unabhängigkeitsannahme auf. Sie schätzt Fähigkeiten von Personen-gruppen, welche sich in naher Nachbarschaft befinden und die Fähigkeiten für die zugehörigen Personen rechtfertigen. Die Beobachtungen werden aber trotzdem weiter als unabhängig voneinander gegebene Antworten behandelt. Bleibt die Anzahl an Fähigkeitsgruppen bei zunehmender Stichprobengröße gleich, steigt die *likelihood* gegen eins, wenn dieser eine eindeutige Spitze zugeordnet ist. Der Messfehler ist für jene Personen gleich, welche dieselbe Personenfähigkeit aufweisen, gleichgültig, welche Antwortmuster diese haben. Wird die ‚richtige‘ Lösung gefunden, weist die richtige Antwortkategorie, bezogen auf die steigende Personenfähigkeit, wiederum einen monoton steigenden Verlauf auf, das unwahrscheinlichste und ‚falsche‘ Item einen monoton sinkenden Verlauf. Die restlichen CCCs weisen keine monotonen Verläufe auf und besitzen ihr Maximum bei einem endlichen Wert. Er empfiehlt cML, aufgrund des Rechenaufwandes der uML.

Auch für longitudinale Untersuchungen gibt es Lösungen. Das *linear logistic latent trait model* wurde entwickelt, um Iteminhalte in die Vorhersage des Itemerfolges zu implementieren. Falls passende Inhaltsfaktoren gefunden

werden, kann auch die Itemschwierigkeit direkt geschätzt werden. Die Messung von Veränderungen ist ein Spezialfall des LLTM. Wird eine Person mit einem Item im Anschluss an eine Beeinflussung die den durchschnittlichen *trait level* verändert konfrontiert, wird eine Konstante hinzugezählt, um die Bedingungen des Items zu messen. Im *multidimensional Rasch model for learning and change* (MRMLC) beinhaltet jede neue Beobachtung eine neue Dimension, auch *Wiener process structure* genannt. Dieses Modell ist für Daten tauglich, welche steigende Varianzen aufgrund der erhöhten Anzahl an involvierten *traits* aufweisen. Der anfängliche *trait level* spielt bei allen Beobachtungen eine Rolle. Der zweite *trait level* repräsentiert die Veränderung von der ersten zur zweiten Beobachtung. Der dritte *trait level* repräsentiert die Veränderung von der zweiten zur dritten Beobachtung, etc. Glück und Spiel (1997) beschreiben anhand von Beispielen vier verschiedene Ansätze zur Messung von Veränderungen mit dichotomen oder ordinalen Daten in Tabelle 32. Die Annahme der Eindimensionalität wird für einige dieser Modelle vernachlässigt. Wird untersucht, ob sich unterschiedliche Gruppen unterschiedlich ändern, wird die Schwierigkeitsveränderung jedes Items einzeln betrachtet. Aussagen können dann nur auf die jeweiligen Gruppen bezogen werden. Dies gilt nicht für das MRM, wo die personenspezifischen Veränderungen gemessen werden.

Modell	Art der Fragestellung	Anwendungsvoraussetzungen
Linear Logistic Test Model (LLTM)	(quasi) experimentell, hypothesenprüfend	Rasch-Homogenität des verwendeten Testmaterials
Linear Logistic Model with Relaxed Assumptions (LLRA)	(quasi) experimentell, hypothesenprüfend	keine
Hybrid-LLRA (HLLRA)	(quasi) experimentell, hypothesenprüfend	Paarweise Rasch-Homogenität der zum 1. und 2. Testzeitpunkt verwendeten Items
Mixed Rasch Model (MRM)	Exploratorisch, Suche nach Klassen von Personen	keine

Tabelle 32: Item Response Modell für die Veränderungsmessung (Glück und Spiel, 1997)

Da das Rasch-Modell von der Eindimensionalität der Spalten ausgeht und die Daten drei Dimensionen aufweisen, Items, Personen und Zeit, müssen diese zur Bearbeitung in eine zweidimensionale Matrix überführt werden. Es gibt zwei Möglichkeiten der Datenumformung. In der ersten werden die Items zu deren Zeitpunkten nacheinander aufgeführt. Dies nennt man die Erzeugung von ‚virtuellen Items‘, da künstlich mehr Items als beobachtet erzeugt werden und

die Datenmatrix in die Breite gestreckt wird. Gilt die Eindimensionalität, werden über die Schätzung der Itemschwierigkeiten Aussagen bezüglich der Personenfähigkeiten getroffen. In der zweiten Umformung werden ‚virtuelle Personen‘ erzeugt, also die Datenmatrix in die Länge gezogen. Jede Person wird für das jeweilige Item zum jeweiligen Zeitpunkt als mehr Personen betrachtet. Hier wird davon ausgegangen, dass jedes Item zu allen Zeitpunkten die gleiche Dimension misst. Die Interpretation der Personenfähigkeiten wird wiederum über die Itemschwierigkeiten interpretiert. Diese Art der Umformung wird später für die Kinostudie im LLRA durchgeführt.

Im LLTM werden Itemschwierigkeiten zu einem späteren Zeitpunkt, mit Hilfe der Itemschwierigkeiten eines vorangehenden Zeitpunktes und ein oder mehr Veränderungsparametern dargestellt. Der Modellvergleich beginnt mit dem quasi saturierten Modell, in welchem für jede Gruppe und jedes Item ein eigener Itemparameter geschätzt wird. In weiteren Schritten kann eine Veränderung über alle Items innerhalb einer Gruppe als gleich angenommen oder auch für alle Gruppen ein Itemparameter definiert werden. Schließlich kann die Entfernung des Veränderungsparameters folgen. Die Modelle werden mit Hilfe des *likelihood ratio tests* (LRT) verglichen, bis das sparsamste Modell aufgefunden wird.

Im LLRA muss die Eindimensionalität nicht gewährleistet sein. Es kann sogar für jedes Item eine eigene Fähigkeit unterstellt werden. Es wird ein eigener Veränderungsparameter für jede Person und jedes Item geschätzt. Besser ist es jedoch die Veränderungen gruppenspezifisch zu schätzen, um stabilere Lösungen zu erhalten. Sind die Veränderungen unterschiedlicher Items in einer Gruppe verschieden, können diese Annahmen als Hypothesen getestet werden. Probleme bezüglich der Instabilität des Modells tauchen dann auf, wenn die Veränderungen aller Items in eine Richtung tendieren. Da im Kinobeispiel sowohl steigende als auch sinkende Veränderungen zu erwarten sind, sollte dies kein Problem darstellen.

Das HLLRA bietet Lösungen für letztgenanntes Problem. Dafür können beispielsweise für zwei Gruppen unterschiedliche Aufgaben, welche paarweise Homogenität aufweisen und raschhomogen sind, zur eindimensionalen Veränderungsmessung für die Itempaare verwendet werden. Die quantitative Veränderung kann jedoch nicht interpretiert werden, lediglich das Vorhandensein eines *treatment*-Effektes, wenn beispielsweise eine der beiden Gruppen eine *treatment*-Gruppe und die andere eine Kontrollgruppe ist. Kann keine Veränderung erkannt werden, entspricht die tatsächliche Veränderung der

Personenfähigkeiten der unterschiedlichen Schwierigkeiten der Items zu den jeweiligen Zeitpunkten. Somit werden sowohl Trendparameter, welche die Fähigkeitsveränderungen für die Gruppen messen, als auch ein *treatment*-Parameter, welcher in der *treatment*-Gruppe den zusätzlichen Einfluss misst, im Modell geschätzt. Der Modellvergleich über LRT beginnt wiederum mit dem quasi saturierten Modell, gefolgt von der Entfernung der Effektparameter, Einführung eines gemeinsamen Trendparameter für alle Items und schließlich die Annahme keiner Veränderung und somit die Entfernung des Trendparameters.

Da in den vorangegangenen Modellen von einem quasi saturierten Modell ausgegangen wird, bei welchem die Beschreibbarkeit der Daten fraglich ist, da unterschiedliche Gruppeneinteilungen unterschiedliche Auswirkungen haben können, werden hier im nächsten Schritt mit Hilfe der MRM explorativ in sich homogene, aber zueinander heterogene Gruppen gesucht. MRM können sowohl für Querschnitts- als auch Längsschnittdaten angewandt werden. Sie verbinden die Rasch Modelle mit den *latent class*-Modellen. Grund dafür ist, dass wenn beispielsweise im gesamten Datensatz das Rasch-Modell nicht zutrifft, es in Subgruppen jedoch zutreffen kann und daher in den Subgruppen unterschiedliche Fähigkeiten erfasst werden. Werden alle Gruppen zusammengezogen, ist dies das klassische Rasch-Modell. Somit können beispielsweise Personengruppen unterschiedlicher kognitiver Entwicklungsstufen ermittelt werden. Es werden zuerst Modelle mit unterschiedlicher Anzahl an Klassen gebildet. Dann wird mit Hilfe eines Vergleiches über AIC, BIC, CAIC oder über die Klassenzuordnungswahrscheinlichkeit, das parameterbezogen schlankere Modell bevorzugt. Für die Klassenbildung gibt es zwei Möglichkeiten. Es können Klassen mit unterschiedlichen Fähigkeitsveränderungen gesucht werden. Aber auch Modelle mit Klassenwechsel sind denkbar. Für diese widerspiegelt die Veränderung der Klassenzugehörigkeit die Veränderung der Personenfähigkeit. Es können quantitative Veränderungen, also Zugehörigkeitswahrscheinlichkeiten interpretiert werden. Diese sind nicht deterministisch, da die Person jener Klasse zugeordnet wird, welche für sie am wahrscheinlichsten zutrifft. Weiters können qualitative Aussagen über die Lösungshäufigkeiten zwischen den Klassen gemacht werden.

Fischer (1989) beschreibt das LLRA für Veränderungsmessungen von dichotomen Daten. Damit werden latente Veränderungen, *treatment*-Effekte und –Interaktionen und Trendeffekte geschätzt. Die Items müssen nicht eindimensional und können unterschiedlich in den jeweiligen Zeitpunkten sein. Den Personen wird keine Populationsverteilung unterstellt. Für jedes Item kann

ein eigener *treatment*-Effekt und ein Trend über alle Items berechnet werden. Für Subgruppen können jeweils wieder einzeln auf die Gruppe zugeschnittene *treatment*-Effekte spezifiziert werden. Ebenso können die Einwirkungen mehrerer *treatment*-Effekte gleichzeitig untersucht werden. Die Modelle werden schließlich mit Hilfe des *conditional likelihood ratio* (CLR)-Tests verglichen, welcher einer asymptotischen χ^2 -Verteilung folgt. Mögliche Modelle sind keine Veränderung, ineffektive oder gleich effektive *treatments*, keine *treatment*-Interaktionen, Generalisierbarkeit der *treatment*-Effekte und des Trends über Personen und Items, oder die Generalisierbarkeit über die Items, so dass alle bzw. Subgruppen von Items bzw. die dahinterliegenden latenten *traits* gleich empfindlich gegenüber dem *treatment*-Effekt sind. Somit ist es oft notwendig die Personen in Subgruppen zu unterteilen, welche ähnlich auf die *treatment*-Effekte reagieren oder die Items in Subgruppen, welche ähnlich empfindlich, den *treatment*-Effekt messen. Dies kommt Interaktionen zwischen *treatment* und Personen bzw. Items gleich.

Das LLRA geht davon aus, dass die gleichen Items zu den jeweiligen Zeitpunkten vorgestellt werden und diese lokal unabhängig sind. Werden aber Personen mit großen Veränderungen beobachtet, ist die Schwierigkeit des Items zu einem späteren Zeitpunkt oft schon zu leicht und es entstehen keine sinnvollen Antwortpatterns mehr. Um dieses Problem zu lösen wird das hybride Modell vorgestellt. Es können dabei unterschiedliche oder auch gleiche Items unterschiedlicher Anzahl zu beliebig vielen Zeitpunkten befragt werden, vorausgesetzt diese messen die selbige latente Dimension. Zumindest ein Item muss den Personen für eine der gemessenen latenten Dimension vorgestellt werden, wobei die Items einer latenten Dimension einem Rasch Modell entsprechen müssen. So kombiniert das LLRA die Multidimensionalitätsannahme, nämlich einen latenten *trait* pro Itempaar mit der Eindimensionalitätsannahme, also die Rasch-Modell Annahme innerhalb der Itempaare. Die Itemschwierigkeiten zu den unterschiedlichen Zeitpunkten werden für die verschiedenen Items mitgeschätzt.

Die Erzeugung der Designmatrix, welche im Kinodatensatz für die Durchführung in R notwendig ist, wird von Fischer (1989) erläutert. Die Daten müssen aus der dreidimensionalen Datenmatrix für den zweidimensionalen Raum aufbereitet werden. Dabei werden über die Zeitpunkte die Items als *virtual items* bezeichnet und die Personen als *virtual subjects*. Das Ergebnis ist eine Blockmatrix, in der Blöcke für die einzelnen *virtual items-virtual subjects*-Kombinationen stehen. Die aufgetretenen Kombinationen werden mit eins gefüllt, die weiteren Kombinationen mit null. Steht in der Matrix eine Eins, wird

für diesen Übergang ein Veränderungseffekt berechnet. Steht eine Null, wird kein Veränderungseffekt berechnet. Im Kinobeispiel wird jede Person als eigener Block betrachtet, da *a priori* keine Subgruppen definiert werden. Tabelle 33 stellt die Blockmatrix für elf Items zu sechs Zeitpunkten dar.

	1	2	3	4	5	6	7	8	9	10	11
1	1										
2		1									
3			1								
4				1							
5					1						
6						1					
7							1				
8								1			
9									1		
10										1	
11											1

Tabelle 33: LLRA-Designmatrix

Um die Auffindbarkeit und einen schnellen Überblick über den Matrixaufbau zu erleichtern, werden leere Zellen nicht mit Nullen gefüllt. Die ersten elf Zeilen beinhalten keine Einsen, da es zum Zeitpunkt eins noch keine Veränderungen gibt. Zeilen zwölf bis 22 stehen für den zweiten Zeitpunkt. Diese elf Zeilen treffen auf die ersten fünf Spalten. Diese fünf Spalten stehen jeweils für die fünf Veränderungsparameter des ersten Items. Wo die Zeile zwölf auf Spalte eins trifft erhält das Datenfeld den Eintrag eins. Es wird also zum Zeitpunkt zwei ein Veränderungsparameter für das Item eins geschätzt. Analog wird dies für alle weiteren Zeitpunkte und Items durchgeführt. Der Originaldatensatz entspricht ursprünglich einer 6x11 Matrix pro Person. Aufgrund der Vervielfältigung der Item- und Personenspalten, spricht man deshalb von einem Aufeinandertreffen von *virtual items* auf *virtual subjects*. Hier werden keine Interaktionseffekte implementiert.

Fischer (1989) bespricht Spezialfälle, welche die Vielfältigkeit dieser Modellgruppe offensichtlich machen. Dazu gehören die Behandlung von mehr als einem Item pro Dimension und Zeitpunkt was über *virtual time points* gelöst wird, die Verwertung von bereits bekannten Itemparametern einer Vorstudie und die wiederholte Vorlage von Items zu unterschiedlichen Zeitpunkten. Das Beispielmodell geht vom saturierten Modell aus, wird zuerst über die Generalisierung der *treatment*-Effekte über die Items eingeschränkt und später werden diese über die Subgruppen gleichgesetzt. Darauf folgt ein Modell, in welchem kurz- und langfristige Effekte unterschieden werden. Es wird lediglich eine kurzfristige Veränderung unterstellt, jedoch keine Veränderung zwischen Zeitpunkt zwei und drei. In einer weiteren Einschränkung werden unterschiedliche *treatment*-Effekte in der ‚knowledge domain‘, aber gleiche für die ‚attitudinal and behavioral domains‘ unterstellt. Im vorletzten Modell wird ein Unterschied zwischen kurz- und langfristigen Effekten lediglich für die ‚knowledge domain‘ angenommen, nicht jedoch für ‚attitudinal and behavioral tendencies‘. Im letzten Modell werden für eine Probandengruppe unterschiedliche kurz- und langfristige Effekte in der ‚knowledge domain‘ unterstellt, für die andere Probandengruppe hingegen werden gleiche kurz- und langfristige Effekte angenommen.

Mair und Hatzinger (2007) liefern eine Einführung in das Paket *eRm* zur Durchführung von IRT Modellen in R. Darin werden Spezifikationen des cML erklärt und für das umfangreichste Modell, nämlich das *linear partial credit model* (LPCM), erläutert. Die Modelle sind hierarchisch genestet, wobei die daraus abgeleiteten Modelle Einschränkungen des LPCM darstellen, wie in Diagramm 40 dargestellt. Sie werden über den gleichen Weg wie das LPCM

bestimmt. Das klassische Rasch-Modell ist ein Spezialfall des RSM. Letztgenanntes wiederum ist ein Spezialfall des PCM.

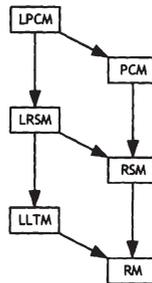


Diagramm 40: Modell Hierarchie (Mair und Hatzinger, 2007)

In der cML Herleitung wird die spezifische Objektivität mathematisch übersetzt. Diese verlangt die Schätzung der Parameter quasi unabhängig von der Stichprobe. Der vorgestellte cML-Ansatz erfüllt diese Bedingungen, nämlich die Separierbarkeit der Item- und Personenparameter. Darunter wird auch verstanden, dass weder den Items noch den Personen eine Verteilung unterstellt wird, was beispielsweise für mML zutrifft. Ob Sampleunabhängigkeit gewährleistet werden kann, wird durch einen graphischen Modellvergleich, bei welchem die Stichprobe gesplittet und die beiden Teile miteinander verglichen werden, überprüft. Das R-Paket bietet aber auch die Möglichkeit *a priori* Gruppen einzubringen und die Teilstichproben über diese Gruppen zu vergleichen. Andersen's LRT überprüft die Raschtauglichkeit der Items. Zuerst wird ein einfaches Rasch-Modell gerechnet. Ist der Test nicht signifikant, sind die Items tauglich und es macht Sinn, dieses Modell mit dem sparsameren *linear logistic test model* (LLTM) zu vergleichen. Aufgrund der geringeren Itemparameteranzahl stellt das LLTM nämlich eine eingeschränkte Version des RM dar.

In einem weiteren Paper erklären Mair und Hatzinger (2007) die Spezifikationen und Eigenheiten weiterer Modelle, das *linear logistic trait model*, LPCM und das *linear rating scale model* (LRSM) mit Ausführungen zu den Spezifikationen virtueller Items. All diese Modelle trennen die Schätzung der Itemparameter und der Personenparameter. Im Package *eRm* in R können über die Designmatrix sowohl Messwiederholungsspezifizierungen als auch Gruppenspezifizierungen eingebracht werden. Eine Darstellung der erforderlichen Matrizen findet sich in Diagramm 41. Diese kann sowohl für die reine Veränderungsmessung, als auch für Veränderungsmessvergleiche zweier Gruppen, nämlich einer *treatment*-Gruppe und einer Kontrollgruppe, verwendet werden. Auch eine Kombination

oder die hypothesengerichtete Schätzung longitudinaler Veränderungen ohne Gruppenspezifisierung sind möglich. Verglichen mit Diagramm 40 ist erkennbar, dass die Zeilen über die Gruppen verdoppelt werden. Auch ein LPCM mit zwei Zeitpunkten und zwei Gruppen, oder die Implementierung spezieller Strukturen, wenn beispielsweise eine Gruppe ein Item beantwortet, aber eine andere nicht und dadurch Fehlerwerte auftauchen, sind möglich.

			η_1	η_2	η_3	η_k	η_{k+1}	η_{k+2}
Time 1	Group 1	$\beta_1^{*(1)}$	1	0	0	0	0	0
		$\beta_2^{*(1)}$	0	1	0	0	0	0
		\vdots			...		\vdots	\vdots
		$\beta_k^{*(1)}$	1	0	0	1	0	0
	Group 2	$\beta_{k+1}^{*(1)}$	1	0	0	0	0	0
		$\beta_{k+2}^{*(1)}$	0	1	0	0	0	0
		\vdots			...			
		$\beta_{2k}^{*(1)}$	1	0	0	1	0	0
Time 2	Group 1	$\beta_1^{*(2)}$	1	0	0	0	1	0
		$\beta_2^{*(2)}$	0	1	0	0	1	0
		\vdots			...		\vdots	\vdots
		$\beta_k^{*(2)}$	1	0	0	1	1	0
	Group 2	$\beta_{k+1}^{*(2)}$	1	0	0	0	1	1
		$\beta_{k+2}^{*(2)}$	0	1	0	0	1	1
		\vdots			...		\vdots	\vdots
		$\beta_{2k}^{*(2)}$	1	0	0	1	1	1

Diagramm 41: Designmatrix wiederholter Messungen mit Subgruppen (Mair und Hatzinger, 2007)

Um zeitliche Veränderungen von Personen zu messen gibt es zwei Ansätze. Entweder die Personenfähigkeitsveränderung wird geschätzt, oder die Personenfähigkeiten werden über die Zeit hinweg konstant gehalten und die gemessenen Veränderungen der Itemschwierigkeiten dienen zur Interpretation der Personenfähigkeitsveränderungen. Letzterer Fall wurde in erwähntem Paket implementiert. Zum ersten Messzeitpunkt bestimmt der Itemparameter die Schwierigkeit. Zum zweiten Zeitpunkt setzt sich die Schwierigkeit aus einer additiven Linearkombination aus Itemschwierigkeit plus *treatment*-Effekt zusammen. Zur Bewertung der Modelle ist ebenfalls ein graphischer Modelltest, nämlich eine Aufspaltung des Gesamtsamples über den *rawscore* der Personen als *default*, als auch der LRT nach Anderson, implementiert.

Hatzinger und Rusch (2009) geben eine Einführung in das R-Paket *eRm* zum Schätzen eines LLRA. *Relaxed assumptions* bedeutet hier, dass weder die Eindimensionalitätsannahme der Items, noch die Verteilungsannahmen der Personenpopulation nötig sind. Im multidimensionalen Ansatz sind so viele Dimensionen möglich, wie es Items gibt. Die Items können zu Dimensionen gruppiert werden, bis diese wiederum in die Eindimensionalität zusammenfallen. Es können sowohl nur zeitliche Veränderungen, als auch Veränderungen aufgrund von zusätzlichen Variablen, wie beispielsweise das Geschlecht, betrachtet werden. Im Kinobeispiel wird für jedes Item eine unterschiedliche Veränderung unterstellt. Im einfachsten Fall gibt es einen gemeinsamen Trend über alle Items. Auch Einflüsse unterschiedlicher *treatments*, oder zeitliche Kovariateneffekte zusätzlich zum zeitlichen Effekt, also dem Trend, sind denkbar. So können beispielsweise unterschiedliche zeitliche *treatment*-Effekte mit unterschiedlichen itemspezifischen Veränderungen geschätzt werden. Im Kinobeispiel werden lediglich zeitlich unterschiedliche Veränderungen für die einzelnen Items zwischen Zeitpunkt eins und Zeitpunkt zwei, Zeitpunkt eins und Zeitpunkt drei, ..., und Zeitpunkt eins und Zeitpunkt zwölf, geschätzt. Auch Modelle mit gleicher oder unterschiedlicher Anzahl an Kategorien, bei welchen die Abstände zwischen den Kategorien gleich sind, können mit dem RSM Ansatz geschätzt werden. Sind die Abstände über die Items zwischen den Kategorien unterschiedlich, kommt der PCM Ansatz zu tragen. Somit ist auch eine zeitliche Veränderung der *thresholds* der einzelnen Items im PCM denkbar, was natürlich die Anzahl an zu schätzenden Parametern drastisch erhöht.

Der *conditional likelihood ratio test* (LQT bzw. LRT) nach Anderson testet die Daten auf Rasch-Konformität. Dafür werden die Daten aufgeteilt und die Teilergebnisse miteinander verglichen. Teilungskriterien können z.B. das Geschlecht sein, oder auch Außenkriterien wie beispielsweise der Mittelwert. Dann werden für die Gruppen die *likelihoods* berechnet. Durch Vergleich wird der χ^2 -Wert berechnet. Ist dieser kleiner als der kritische χ^2 -Wert wird die Nullhypothese beibehalten. Die Nullhypothese steht für die Aussage, dass das Rasch-Modell gilt. Ist der Wert größer wird die Raschkonformität verworfen. Im Kinobeispiel wird als externes Teilungskriterium der Median verwendet. Es muss jedoch beachtet werden, dass ein nicht signifikanter Test noch keine Bestätigung für Raschkonformität ist. Die Daten müssen weiter anhand von externen oder internen Teilungskriterien getestet werden. Dadurch kann zumindest ein Ausschluss für das Nichtgelten eines Gruppenteilungskriteriums geschaffen werden. Abgesehen vom Gesamttest der Items mit Hilfe des LRT und dem graphischen Modelltest können die Items auch mit einem Wald-Test überprüft werden. Für die Kinomodelle wird als Teilungskriterium der Median

verwendet. Weiterführende Interpretationen zu den Teilungskriterien verhalten sich analog zum conditional LRT. Der Wald-Test gibt z-Werte und Irrtumswahrscheinlichkeiten aus. Liegt letztere über 0,05, gibt es kein signifikantes Item und es werden alle Items im Modell beibehalten. Auch der Itemfit zur Tauglichkeit der Items kann bestimmt werden. Ist dieser nicht signifikant, gilt für das Item Raschkonformität. Selbiges ist für die Personenparameter möglich. Um Modellvergleiche durchzuführen, werden AIC, BIC oder CAIC verwendet. Das Modell mit dem kleineren Wert wird bevorzugt.

Um die Rechenzeit zu verkürzen, wird für die longitudinalen Untersuchungen kein LPCM berechnet, sondern das dichotome LLRA. Die Personenanzahl spielt bei der Dauer der Berechnung laut Simulationsstudien keine wesentliche Rolle. Einfluss auf die Rechenzeit nimmt vor allem die Anzahl der Items und Gruppen. Gruppen gibt es in den Kinodaten keine, aber elf Items. Somit wird das einfache dichotome Rasch Modell verwendet.

Ostini und Nering (2006) sehen die Vorteile der IRT in der Möglichkeit, die *Item-location* und den Personenlevel auf einer Skala abbilden zu können. Der monotone Verlauf wird auf der *item response function* (IRF) dargestellt und benötigt zwei Parameter die geschätzt werden müssen. Im *logistic ogive model* wird die *location* immer durch das Zentrum der Funktion, nämlich die Mitte zwischen oberer und unterer Asymptote, bestimmt. Der zweite Parameter bestimmt die Steigung der Kurve. Der *inflection point* stellt die steilste Stelle des Kurvenverlaufes dar.

Um zu den Itemkategoriewerten zu gelangen, gibt es verschiedene Wege der Dichotomisierung der Items. Es ergibt sich eine Vielzahl an Möglichkeiten für polytome Modelle. Antworten auf Likertskalenitems werden auch *graded responses* genannt. In derartigen polytomen Modellen gibt es zwei grundlegende Unterschiede zur Bestimmung der Kategorien bzw. Grenzen über die bedingten Wahrscheinlichkeiten. Die erste beschreibt die Wahrscheinlichkeit innerhalb einer Kategorie zu antworten, die zweite die Wahrscheinlichkeit positiv und nicht negativ auf einer Grenze zwischen zwei Kategorien zu antworten. Verglichen mit dem dichotomen Modell, wo diese beiden Möglichkeiten zur selben Lösung gelangen, gibt es im polytomen Modell immer zumindest eine Kategorie, welche durch zwei Grenzen bestimmt wird. Durch Verbindung der dichotomisierten Grenzen wird die Wahrscheinlichkeit einer Antwort innerhalb einer Kategorie zu antworten bestimmt. Polytome Modelle werden durch die Wahrscheinlichkeit auf einer Kategoriengrenze zu antworten bestimmt. Dies wird *category boundary response function* (CBRF) genannt und bestimmt zum

speziellen *trait* die Hälfte der Antwortwahrscheinlichkeit. Die Zweite Möglichkeit wird durch die Wahrscheinlichkeit innerhalb einer Kategorie zu antworten durch die *item category response function* (ICRF) bestimmt.

Dem Prinzip der spezifischen Objektivität zugrundeliegend, dürfen Itemparameter nicht von Personenparametern und anderen Itemparametern abhängen. Bei der Bestimmung der IRF im dichotomen Modell wird dafür der Diskriminierungsparameter nicht geschätzt, sondern nur der *location*-Parameter. Somit entstehen zwei grundlegende Unterschiede. Im polytomen Rasch Modell wird das dichotome Rasch-Modell mit der IRF zur Bestimmung der Kategoriengrenzen verwendet. Im polytomen Thurstone/Samejima-Modell wird das *two-parameter logistic* (2PL) Modell für die IRF zur Bestimmung verwendet. Das Rasch-Modell behält die spezifische Objektivität durch Dichotomisierung, die CBRF durch die Wahrscheinlichkeit positiv auf einer Kategoriengrenze zu antworten, bei. Die Möglichkeiten werden in Diagramm 42 dargestellt. Die Dichotomisierung kann nur die benachbarten Kategorien der Grenze, unterer Teil der Graphik, oder alle oberhalb und unterhalb liegenden Kategorien, oberer Teil der Graphik, mit einbeziehen. Auch eine Kombination beider Möglichkeiten im hybriden Modell ist denkbar.

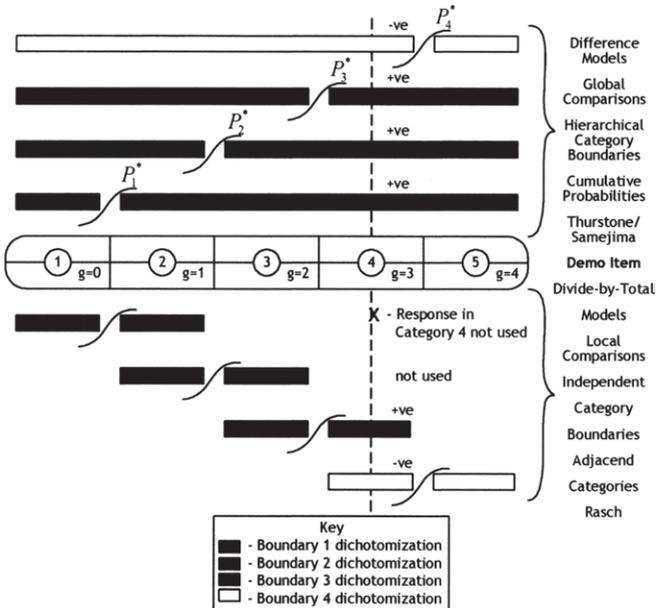


Diagramm 42: Bestimmung polytomer IRT-Modelle (Ostini und Nering, 2006)

Rasch Modelle bewahren die suffiziente Statistik und spezifische Objektivität und beziehen lediglich die benachbarten Kategoriegrenzen mit ein. Sie ignorieren die Antworten in den weiteren Kategorien. Im Thurstone/Samejima-Modell werden alle oberhalb und unterhalb liegenden Kategorien miteinbezogen und betrachten somit die Kategorienantworten in einem gesamten Itemkontext. Die Rasch-Modell CBRF modelliert die Antworten für zwei Kategorien, die Thurstone/Samejima CBRF für jede Kategorie.

Für einen Modellvergleich erwähnen sie den Vergleich der *response functions* zwischen beobachteten und im Modell vorhergesagten oder die graphische Bewertung der Residuen. Der residuenbasierte Ansatz vergleicht die beobachteten mit den vorhergesagten *response functions*. Der multinomiale verteilungsbasierte Test vergleicht das gefittete Modell mit dem allgemeinen multinomialen Modell indem es die beobachteten und erwarteten Häufigkeiten an Antwort-*patterns* vergleicht. Da die gemeinsame Verteilung aller Antwortpatterns im Falle eines guten Fits die gleiche *likelihood* wie die multinomiale Verteilung, die asymptotische χ^2 -Verteilung, besitzt, stehen mehrere Tests zur Verfügung, *Person's- χ^2* , *log-likelihood ratio statistic (χ^2)*, *Freeman-Tukey statistic* und die *Neyman-modified- χ^2* . Die *response function*-basierten Tests, welche die Personenfits betrachten, vergleichen die beobachteten und die erwarteten *log-likelihoods* der individuellen Itemantworten. Der *Guttman error-based* Test als nicht parametrischer Test betrachtet die Zählung der Fehler der *Guttman response errors* über die Itempaare.

Mazanec wendet bereits im Jahre 1979 IRT-basierte Modelle auf ein Beispiel des Imagetransfers an und untersucht, ob das Image einer Zigarettenmarke auf Produkte wie Parfum, Bier, Weinbrand, Bohnenkaffee, Mineralwasser oder Rasierwasser übertragen werden kann. Die latente Variable, welche dabei ermittelt wird, stellt auf Personenseite den Parameter für die Transferbereitschaft dar und auf Itemseite den für die Transferleichtigkeit. Eigenschaften, welche raschkonformen Daten unterstellt werden, werden für den Anwendungsfall nicht bestätigt. Er beschreibt die spezifische Objektivität treffend: ‚Einfach gesagt bedeutet es, dass die Messungen innerhalb von Personenstichproben nicht durch individuelle Antworttendenzen gestört werden können. Die klassische Messtheorie vermochte ja das Dilemma nicht zu umgehen, dass jede in eine Stichprobe neu aufgenommene Person stets auch eine ungewisse Portion individueller Unschärfe einbringt.‘

Eine weitere Studie von Ganglmair und Lawson (2003) stellt eine Skalenentwicklungsmöglichkeit vor, welche ein klassisches Problem sozialempirischer

Untersuchungen lösen soll. Zufriedenheitsanalysen wurden früher dem kognitiven Bereich zugeschrieben. Im Laufe der Zeit wurde diese immer mehr mit dem affektiven Bereich verknüpft. Die aktuell gebräuchlichen Skalen der CTT, um Zufriedenheit zu messen, weisen geringe Diskriminierung und starke negative Schiefe auf, da selbstbeurteilte Zufriedenheitsverteilungen in die Richtung ‚zufrieden‘ tendieren und somit die meisten Antworten auch die positivsten darstellen. Ausgehend von der *affective response to consumption scale* (ARC) werden in genannter Studie Items aus der *delighted-terrible* (D-T) *scale* und anderen verwendet, um auch Personen in den sehr hohen Zufriedenheitslagen zu diskriminieren. Auf einer ‚Ja/Nein‘-Skala wurde angekreuzt ob das Adjektiv zutrifft. Da die negativ formulierten Items zu selten gewählt wurden, wurden diese ausgeschlossen. Der *person separation index* wurde verwendet um interne Konsistenz zu messen. Dieser liegt zwischen null und eins und misst den Anteil wahrer zu beobachteter Varianz, basierend auf den Schätzungen der Personenfähigkeit. Die Eindimensionalität der ARC Skala wird durch χ^2 -Test des Fits der beobachteten Items mit den erwarteten gemessen. Dieser war signifikant und wurde als akzeptabel betrachtet. Die Entfernung eines Items aus der Skala führte nicht zu einer großen Beeinträchtigung der Eindimensionalität und Reliabilität, wenn dieses eine χ^2 -Wahrscheinlichkeit unter 0,01 aufwies. Externe Variable hatten keinen signifikanten Einfluss. Die neu entwickelte Skala wurde mit Hilfe einer geteilten Stichprobe mit der D-T-Skala und zwei 7-Punkt-Likert Skalen verglichen. Eine ANOVA konnte die bessere Diskriminierung der neuen Skala in hohen Zufriedenheitsbereichen zeigen.

Ganglmair-Wooliscroft (2007) untersuchte anhand von zwei Fallstudien die Raschtauglichkeit der ARC Skala. Die Ergebnisse einer Mobiltelefonstudie und die einer Zugreise wurden miteinander verglichen. Zuvor wurden Items, welche die Raschkonformität nicht erfüllen und somit der χ^2 -Wert unter 0,05 lag, entfernt, damit das beobachtete dem erwarteten Modell zu Genüge entspricht. Es finden sich Unterschiede bezüglich der Lage der Items in beiden Studien. Die Befürwortung emotionaler Items findet sich in beiden Studien in unterschiedlicher Anordnung. Diese Unterschiede können aus dem Situationskontext der Befragung entstehen. Die Verwendung einer Vielzahl hoher positiver Statements hilft die hohe positive Schräglage von Zufriedenheitsmessungen zu umgehen und steigert die Diskriminierung zwischen den Respondenten. Das Gesamtmodell wurde durch einen signifikanten χ^2 -Wert gerechtfertigt. Zusätzlich wurden die Residuen untersucht. Mit diesen wurde eine Hauptkomponentenanalyse durchgeführt. Erklärt ein Faktor einen hohen Anteil der

Varianz, taucht die Vermutung auf, dass auch eine andere Dimension auf die Items Einfluss hat. Dies traf auf die angesprochene Studie nicht zu.

4.3.2.1. *Rasch Model - Kinodaten*

Im Rahmen der Kinostudie werden alle Zufriedenheitsbewertungen sämtlicher Kinobesuche der ersten sechs Wellen miteinbezogen. Für das erste Modell wird die 6-Punkt-Likert-Skala dichotomisiert. Die Antworten eins, zwei und drei, welche die Bewertungen zufriedener Kunden darstellen, erhalten den Wert null. Die Antworten vier, fünf und sechs, welche die Bewertungen unzufriedener Kunden darstellen, erhalten den Wert eins. Für die vorliegende Studie werden die Parameter der Itemschwierigkeit durch ‚Servicequalitätslevel‘ und Personenfähigkeit durch ‚Erwartungserfüllung‘ ersetzt. Es wird bestimmt, ob es für die Items der Zufriedenheitsmessungen schwer oder leicht ist, zufriedene bzw. unzufriedene Antworten zu geben. Auch werden manche Items die befragten Personen besser als andere differenzieren, da diese genau auf jener kritischen Zufriedenheitsposition liegen, auf welcher sich der Großteil der Personen gerade befindet. Es ist für diese auch am schwersten, zwischen positiver und negativer Zufriedenheit zu wählen. Es wird ein einfaches Rasch-Modell gerechnet. Die *location* Parameter in Tabelle 34 beschreiben die Lage der Items und somit die Leichtigkeit bzw. Schwierigkeit, positive bzw. negative Antworten zu geben. Die 95%-igen Konfidenzintervalle mit zugehöriger Standardabweichung geben den Bereich der *location* an, in welchem der durchschnittliche Übergang zwischen positiver und negativer Bewertung zu erwarten ist. Dieser Übergang liegt an jenem Punkt, wo eine 50%-ige Wahrscheinlichkeit positiv zu antworten und eine 50%-ige Wahrscheinlichkeit negativ zu antworten besteht. *Location* ist die Lage auf der aktuellen Zufriedenheitsskala, eine latente Dimension, welche die Bewertung auf einer allgemeinen Zufriedenheitsempfindungsskala zulässt.

Jene Items, welche sehr niedrige *locations* aufweisen, z.B. Technik/Bild/Ton, werden eher bessere Zufriedenheitsbewertungen erhalten, da der Übergang von negativ auf positiv schon sehr früh stattfindet. Jene mit sehr hohen Werten, z.B. Buffet, werden eher schlechte Zufriedenheitsurteile erhalten, da ein verbessertes Empfinden der Zufriedenheit benötigt wird, um den Übergang einer negativen zu einer positiven Bewertung zu erreichen. Es gibt für alle *locations* Wahrscheinlichkeiten für eine Bewertung auf der Spanne zwischen ‚sehr zufrieden (0)‘ und ‚nicht zufrieden (1)‘. Angenommen, man möchte für eine Person die Wahrscheinlichkeit dafür bestimmen, mit welchem Wert höchstwahrscheinlich auf die einzelnen Items geantwortet wird, so muss lediglich eine vertikale Linie durch den Personenwert in Diagramm 43 gezogen werden, in diesem Fall durch

den Wert auf der Erwartungserfüllungsdimension. Die Itemnummerierung sämtlicher Graphiken dieses Kapitels folgt der Auflistung aus Tabelle 34. Die Wahrscheinlichkeiten, negative Zufriedenheitsbewertungen für Technik/Bild/Ton abzugeben, liegt für eine Person, welche sich auf einem Zufriedenheitslevel von null auf der x-Achse befindet, unter 50 %, für das Item Buffet jedoch über 50 %. Dies bedeutet, dass die Person Technik/Bild/Ton eher positiv und Buffet eher negativ bewerten wird. Je weiter ein Item links angesiedelt ist, desto größer die Wahrscheinlichkeit negativer Antworten, je weiter rechts, desto größer die Wahrscheinlichkeit positiver Antworten.

Item	Location	Standardabweichung	Unteres/Oberes Konfidenzintervall
Technik/Bild/Ton (1)	-0.913	0.041	-0.992/-0.833
Komfort (2)	-0.414	0.034	-0.481/-0.347
Servicepersonal (3)	-0.071	0.032	-0.134/-0.007
Preis/Leistung (4)	0.880	0.025	0.831/0.929
Buffet (5)	1.445	0.027	1.392/1.497
Atmosphäre (6)	0.327	0.028	0.273/0.381
Standort/Erreichbarkeit (7)	-0.351	0.033	-0.416/-0.285
Filmangebot (8)	-0.417	0.035	-0.485/-0.349
Informationsangebot (9)	-0.265	0.035	-0.333/-0.197
Reservierungsmöglichkeiten (10)	-0.216	0.038	-0.289/-0.142
Erscheinungsbild/Image (11)	-0.005	0.030	-0.065/0.054

Tabelle 34: Locations - dichotomes Modell (123-456)

Im diesem Modell wird davon ausgegangen, dass alle Items die gleiche Steigung besitzen. Das heißt, der Übergang von ‚sehr zufrieden‘ bis ‚nicht zufrieden‘ läuft bei allen Items in einem gleichbleibenden Stärkenänderungsintervall ab. Die einzelnen elf Kurven können sich in aufgrund deren identischer Steigungen nicht überschneiden. Werden alle Kurven steiler geschätzt, passiert der Übergang schneller und eine kleine Veränderung der angebotenen Serviceleistung mit der einhergehenden Zufriedenheitsbewertung wird zu einem schnelleren Schwanken zwischen einer positiven und negativen Antwort führen. Je flacher die Kurve, desto mehr Einwirkung und Veränderung wird benötigt, ein Kippen der Antwort von positiv auf negativ zu bewirken.

Joint ICC Plot

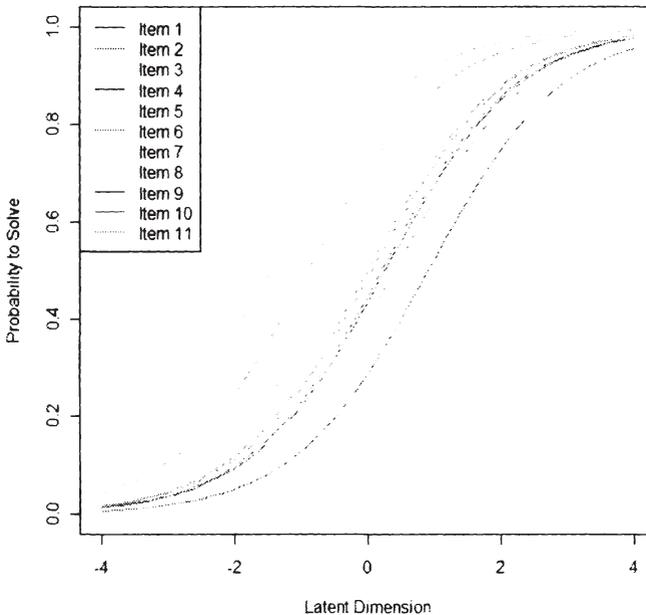


Diagramm 43: Joint-Plot der ICC-Kurven - dichotomes Modell (123-456)

Die x-Achse kann für die Leichtigkeit einer gewissen Antwort stehen und von rechts nach links gelesen werden. Je weiter ein Item rechts liegt, desto leichter fällt es positive, je weiter links, desto leichter negative Antworten zu geben. Die x-Achse kann aber auch für die Schwierigkeit einer gewissen Antwort stehen. Je weiter ein Item links liegt, desto schwerer fällt es positive, je weiter rechts, desto schwerer negative Antworten zu geben. Zusammengefasst bedeutet dies, dass bei einer größeren Wahrscheinlichkeit eher eine negative Antwort gegeben wird und bei einer niedrigen eher eine positive. Es ist auch anhand der *location* Parameter aus Tabelle 34 erkennbar, dass Buffet an der linken Seite ausgerichtet ist, also beim Wert null auf der x-Achse eine höhere Wahrscheinlichkeit besitzt und somit den Wert eins, nicht zufrieden, schon sehr früh erreicht. Umgekehrt ist erkennbar, dass Technik/Bild/Ton an der rechten Seite ausgerichtet ist, also beim Wert null auf der x-Achse eine niedrigere Wahrscheinlichkeit besitzt und somit den Wert eins, nicht zufrieden, später als Buffet erreicht.

Aus Vergleichszwecken wird die Lösung veranschaulicht, wenn die Antworten eins und zwei auf null, und die Werte drei, vier, fünf und sechs auf eins,

recodiert werden. Diese Lösung wird aufgrund der Positivlastigkeit der Antwortskalen in sozialempririschen Studien angeführt und stellt hier eine abgeänderte Form der Dichotomisierung dar. Das Ergebnis wird in Diagramm 44 dargestellt.

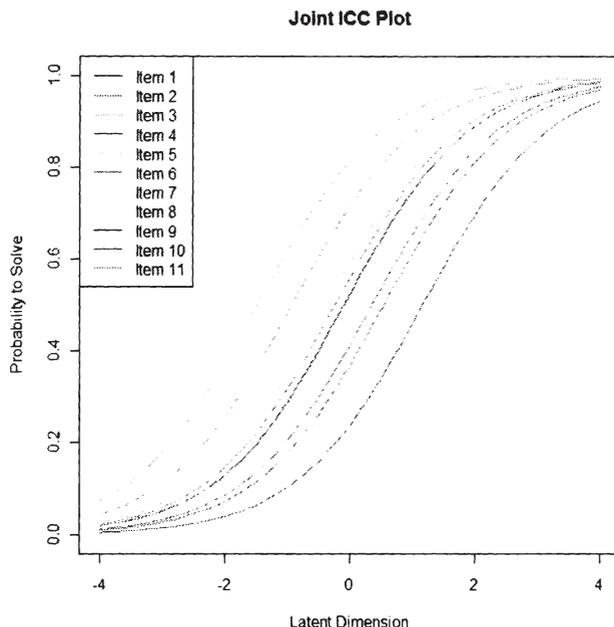


Diagramm 44: Joint-Plot der ICC-Kurven - dichotomes Modell (12-3456)

Die geänderte dichotome Kodierung der Antwortskala bewirkt ein leichtes Auseinanderdriften der Items. Dies resultiert aus der vermehrten Häufigkeit negativer Antworten. Auch in Tabelle 35 ist der Unterschied erkennbar. Die Spanne zwischen dem höchsten und dem niedrigsten *location*-Wert vergrößert sich von 2,358, -0.913 bis 1.445, im 123-456-Modell auf 2,646, -1.169 bis 1.477, im 12-3456-Modell. Es ergibt sich eine Differenz von 0,288.

4.3.2.2. Modellevaluierung - Kinodaten

Aufgrund einer besseren Erklärbarkeit des 123-456-Modells werden die folgenden Tests auf Basis dieser Skalierung vorgenommen. Zuerst prüft ein *conditional* LRT nach Andersen die Raschkonformität der Items. Der Datensatz wird beim Median in zwei Gruppen geteilt. Ein LR-Wert von 58,101 und 10 Freiheitsgraden ergibt einen signifikanten Wert. Die beiden Gruppen sind signifikant voneinander unterschiedlich und die Datenstruktur ist nicht

raschkonform. Tabelle 36 zeigt die Unterschiede beider Gruppen anhand der Parameterwerte.

Item	Location	Standardabweichung	Unteres/Oberes Konfidenzintervall
Technik/Bild/Ton	-1.169	0.025	-1.218/-1.120
Komfort	-0.359	0.020	-0.399/-0.319
Servicepersonal	0.232	0.019	0.194/0.270
Preis/Leistung	0.922	0.018	0.887/0.957
Buffet	1.477	0.021	1.437/1.518
Atmosphäre	0.371	0.018	0.335/0.407
Standort/Erreichbarkeit	-0.636	0.022	-0.678/-0.593
Filmangebot	-0.464	0.021	-0.506/-0.593
Informationsangebot	0.095	0.020	0.055/0.134
Reservierungsmöglichkeiten	-0.545	0.025	-0.593/-0.497
Erscheinungsbild/Image	0.076	0.019	0.039/0.114

Tabelle 35: Locations - dichotomes Modell (12-3456)

Items	Gruppe 1	Gruppe 2
Technik/Bild/Ton	-0.640706393	-0.91873214
Komfort	-0.526484563	-0.40843663
Servicepersonal	-0.365204469	-0.06222220
Preis/Leistung	0.785625438	0.88651508
Buffet	1.524657261	1.44300843
Atmosphäre	-0.034468395	0.34096393
Standort/Erreichbarkeit	0.625623987	-0.39687821
Filmangebot	-0.483775772	-0.41306004
Informationsangebot	-0.391822916	-0.26123819
Reservierungsmöglichkeiten	-0.006329998	-0.21776408
Erscheinungsbild/Image	-0.487114180	0.00784404

Tabelle 36: Likelihood-Ratio-Test nach Andersen für das 123-456-Modell

Diagramm 45 überprüft graphisch, welche Items die stärksten Abweichungen aufweisen. Je weiter sie von der Diagonalen entfernt sind, desto größer die Unterschiede in den gebildeten Gruppen. Die größten Abweichungen zeigt Standort/Erreichbarkeit.

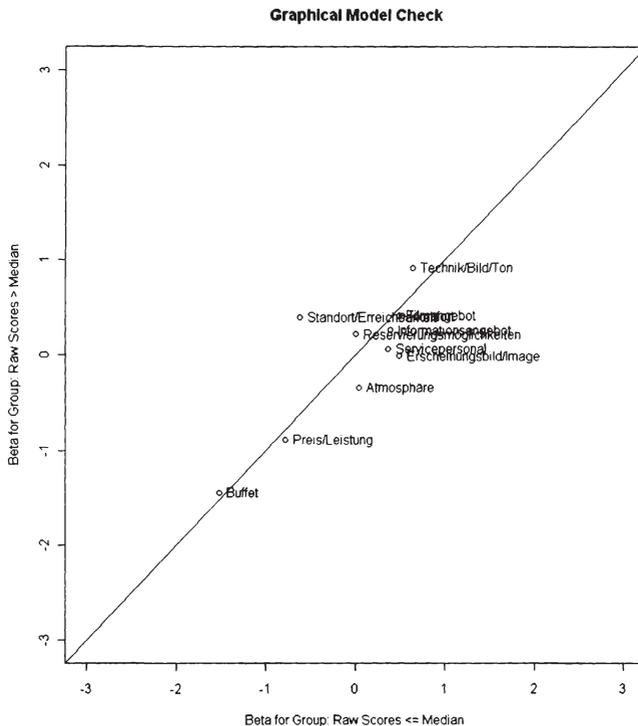


Diagramm 45: Graphischer Modelltest (123-456)

Mit Hilfe eines Wald-Tests in Tabelle 37 werden aufgrund der Trennung nach dem Median und einer zusätzlichen zufälligen Trennung z-Werte und die dazugehörigen Wahrscheinlichkeiten ausgegeben. Fällt einer der Werte unter das gewählte Signifikanzniveau, hier 95 %, finden sich signifikante Unterschiede zwischen den Gruppen. Standort/Erreichbarkeit ist hoch signifikant und auch die Werte von Erscheinungsbild/Image und Atmosphäre sind signifikant. Auch beim graphischen Modelltest zählten diese zu den größten Ausreißern. Diese Items sind nicht raschkonform. Um in den weiteren Analysen jedoch Erkenntnisse und Unterschiede veranschaulichen zu können, werden die nicht-raschkonformen Items nicht aus den Modellen entfernt. Gerade diese Items geben Aufschluss über inhärente Unterschiede und Eigenheiten der Items. Auf die Unterschiede in den Bewertungen von Standort/Erreichbarkeit wird noch genauer eingegangen. Es widerspiegelt schließlich die Lage des Kinos und eine differenzierte Bewertung der Besucher erscheint nicht verwundernswert.

Items	z-Statistik	p-Wert
Technik/Bild/Ton	1.192	0.233
Komfort	-0.555	0.579
Servicepersonal	-1.283	0.199
Preis/Leistung	-0.693	0.488
Buffet	0.668	0.504
Atmosphäre	-2.119	0.034
Standort/Erreichbarkeit	7.216	0.000
Filmangebot	-0.302	0.763
Informationsangebot	-0.450	0.653
Reservierungsmöglichkeiten	0.772	0.440
Erscheinungsbild/Image	-2.172	0.030

Tabelle 37: Wald-Test (123-456)

4.3.2.3. Rating Scale Model - Kinodaten

Vorangegangene Modelle wurden unter der Restriktion dichotomisierter Daten durchgeführt. Nun werden die Daten unter ihrer originalen Beantwortungsskala betrachtet. Die Analysen werden auf Basis der polytomen 6-Punkt-Likert-Skala vorgenommen. Es werden für jedes Item fünf *thresholds* mitgeschätzt. Diese stellen die fünf Kategorienübergänge zwischen eins und zwei, zwei und drei, drei und vier, vier und fünf, und fünf und sechs auf der Zufriedenheitsskala dar. Je nachdem ob der Übergang zweier Bewertungen mehr oder weniger Einwirkung bedarf, können unterschiedliche Modelle formuliert werden. Zuerst werden die Items mit gleichen Steigungen, also gleichen Abständen der Übergänge, geschätzt. Die Abstände zwischen eins und zwei, zwei und drei, drei und vier, vier und fünf, und fünf und sechs, dürfen jedoch unterschiedlich groß sein. Zwischen den Items dürfen somit lediglich die *locations* variieren. Tabelle 38 zeigt die Ergebnisse des RSM.

Je höher (niedriger) die *locations*, desto leichter (schwerer) ist es positive Zufriedenheitswerte zu geben bzw. je höher (niedriger) die *locations*, desto schwieriger (leichter) ist es negative Zufriedenheitsbewertungen zu geben. Die Differenzen zwischen *threshold* eins und zwei betragen 1,28292, zwischen zwei und drei 0,80526, zwischen drei und vier 0,14162 und zwischen vier und fünf -0,36604, für alle elf Items. Je höher die *threshold*-Differenz zweier *thresholds*, desto stärker muss die Zufriedenheitseinwirkung sein, um den Übergang zu einer anderen Bewertung zu bewirken. Theoretisch müssten diese durchgehend abfallen bzw. ansteigen um eine logische Tendenz zwischen Leichtigkeit bzw. Schwierigkeit aufzuzeigen. Die Differenz zwischen *threshold* vier und fünf ist

hier jedoch negativ. Der Übergang zwischen einer Bewertung auf der originalen Zufriedenheitsskala zwischen eins und zwei bedeutet, dass es schwieriger ist, auf das Item mit eins zu antworten. Eine negative Differenz zwischen den *thresholds* macht für diese Erklärung wenig Sinn. Es ist nämlich schwieriger die Zufriedenheit mit sechs zu beantworten, als mit fünf, obwohl sie auf einem höhern Zufriedenheitslevel auf der x-Achse liegt. Diese Abweichung resultiert aus der seltenen Auswahl der negativen Zufriedenheitsbewertungen. Wären die Itemkategorien ordinal angeordnet, gäbe es für jede Antwortkategorie im Vergleich zu den weiteren fünf Antwortkategorien eine Stelle, an welcher die Wahrscheinlichkeit diese Itemkategorie zu wählen am höchsten ist.

Item	Location	Thres- hold 1	Thres- hold 2	Thres- hold 3	Thres- hold 4	Thres- hold 5
Technik/Bild/Ton	2.10331	0.61038	1.89330	2.69856	2.84018	2.47414
Komfort	1.65228	0.15935	1.44226	2.24753	2.38915	2.02311
Servicepersonal	1.34019	-0.15274	1.13017	1.93543	2.07706	1.71102
Preis/Leistung	1.00821	-0.48472	0.79819	1.60345	1.74507	1.37904
Buffet	0.71990	-0.77303	0.50988	1.31514	1.45676	1.09073
Atmosphäre	-0.23900	1.25393	1.04391	1.84917	1.99080	1.62476
Standort/ Erreichbarkeit	1.93191	0.43898	1.72189	2.52716	2.66878	2.30274
Filmangebot	1.78145	0.28852	1.57143	2.37669	2.51831	2.15228
Informations- angebot	-0.05369	1.43924	1.22922	2.03448	2.17610	1.81007
Reservierungs- möglichkeiten	1.82599	0.33306	1.61598	2.42124	2.56286	2.19683
Erscheinungsbild/ Image	-0.12710	1.36583	1.15581	1.96107	2.10270	1.73666

Tabelle 38: Locations und Thresholds - RSM

Um die RSM Unterschiede zweier Items zu visualisieren, werden in Diagramm 46 die beiden am weitesten auseinander liegenden Items abgebildet. Das Kurvengebilde des Standortes liegt sehr weit rechts. Im Gegensatz dazu liegt das identische Kurvengebilde von Buffet weiter links. Es fällt den Probanden schwerer positive Zufriedenheitsantworten in den Kategorien null und eins für Buffet zu vergeben.

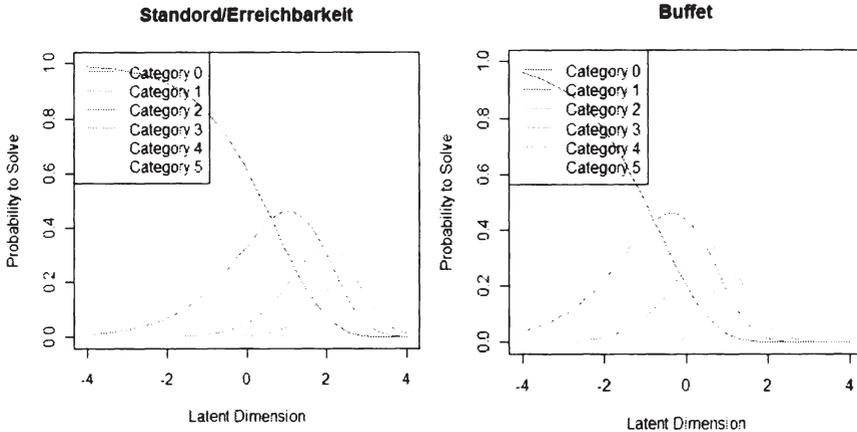


Diagramm 46: Standort vs. Buffet im RSM

4.3.2.4. *Partial Credit Model* - Kinodaten

Nun stellt sich jedoch die Frage, ob die Annahme gleichbleibender *threshold*-Differenzen über die Items rechtfertigbar ist. Vor Durchführung des PCM, welches diese Erweiterung im Gegensatz zum RSM beinhaltet, wird hier noch der Übergang zwischen den unterschiedlichen Antwortkategorien, anhand der Abstände zwischen den einzelnen *thresholds* der elf Items, im Diagramm 47 dargestellt. Je weiter die Kurven voneinander entfernt liegen, desto mehr Zufriedenheitseinfluss bedarf es, einen Übergang zwischen den Antwortkategorien zu bewirken, je näher diese zusammenliegen, desto weniger. Allgemein ist aufgrund der Anordnung der Kurven wiederum erkennbar, dass die Wahrscheinlichkeiten auf der Zufriedenheitsskala nicht ordinal-skaliert angeordnet werden können. Die *threshold*-Übergangskurven von Buffet sind im Vergleich zu Standort/Erreichbarkeit sehr weit verteilt. Dies bedeutet, dass beim Standort der Übergang zwischen ‚sehr zufrieden‘ und ‚nicht zufrieden‘ sehr schnell passiert. Daraus folgt, dass der Standort kurz gesagt entweder zufriedenstellend ist, oder nicht. Bewertungen dazwischen gibt es kaum. Im Gegensatz dazu liegen beim Buffet die Kurven sehr weit auseinander. Daraus lässt sich schlussfolgern, dass bei diesem Item sehr wohl unterschiedliche Zufriedenheitsbewertungen auftauchen. Variationen im angebotenen Servicekriterium bieten einen langsameren breiteren Übergang zwischen sehr und nicht zufrieden.

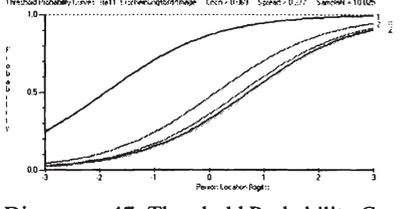
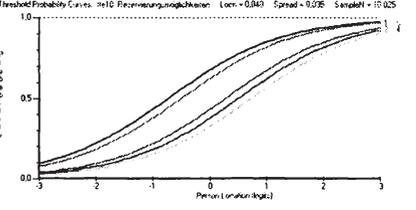
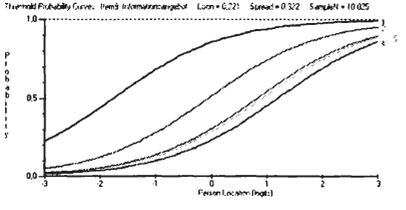
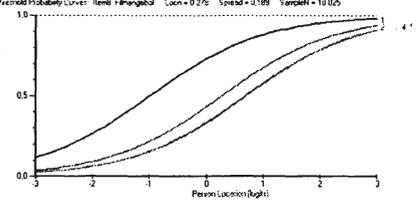
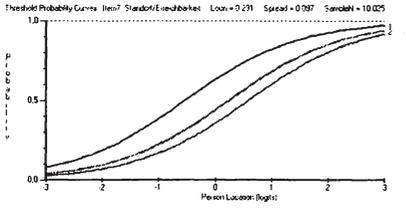
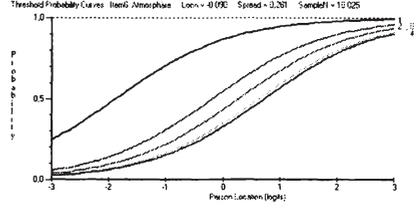
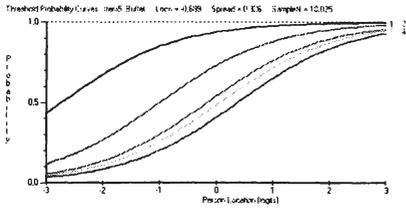
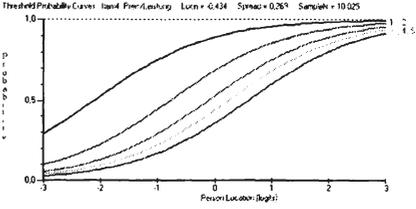
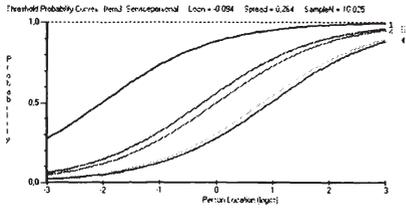
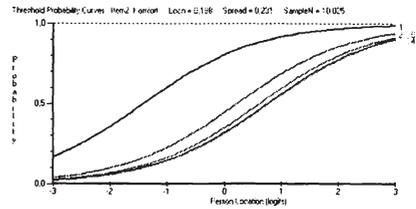
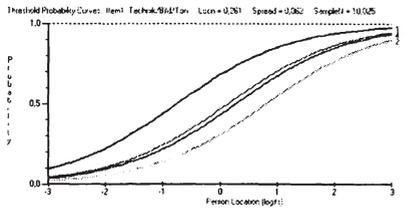


Diagramm 47: Threshold Probability Curves

Diagramm 48 verdeutlicht die beschriebenen Zusammenhänge anhand des Buffets. Auf einer Wahrscheinlichkeitshöhe von 50 % auf der y-Achse befindet sich eine Trennlinie. Diese beschreibt den Übergang zwischen Antwortkategorie eins und zwei, zwei und drei, drei und vier, vier und fünf, und fünf und sechs. Dort wo die strichlierten Kurven diese Grade schneiden, befindet sich der *threshold*, also der Übergang zwischen einer Kategorie in die nächste. Die durchgehenden Linien beschreiben wiederum die Wahrscheinlichkeiten für die einzelnen Antwortkategorien. Sie summieren sich auf jedem Zufriedenheitslevel auf 100 % auf. Je nachdem auf welchem Zufriedenheitslevel sich eine Person befindet, fällt die höchste Wahrscheinlichkeit die Entscheidung, in welchem Übergang zwischen zwei *thresholds* sich eine Person gerade am wahrscheinlichsten befindet. Personen auf der linken Seite der Graphik tendieren eher dazu, in den höheren Zufriedenheitsbereichen zu antworten. Personen auf der rechten Seite tendieren eher dazu, nicht zufriedene Bewertungen abzugeben.

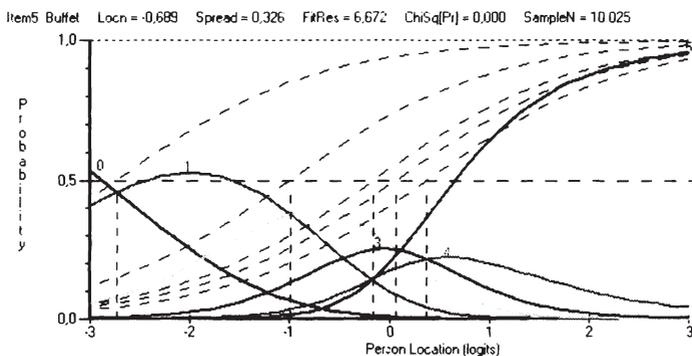


Diagramm 48: Wahrscheinlichkeitsübergänge über die ICCs

In Diagramm 49 werden die Übergangskurven dargestellt. Auf der y-Achse befindet sich der erwartete Wert der Antwort. Für Item 5 Buffet, fallen die Zufriedenheitsbewertungen im Vergleich zu den weiteren Items schlecht aus. Anhand der Steigung der Kurven wird zusätzlich der Übergang visualisiert. Die Kurven können sich aufgrund der unterschiedlichen Steigungen überschneiden. Flache Kurven wie beim Buffet ermöglichen großen Spielraum für unterschiedliche Zufriedenheitsbewertungen. Für Items mit steileren Kurven spielen die mittleren Zufriedenheitsstufen kaum eine Rolle. Es wird empfohlen die 6-Punkt-Likert-Skala auf eine schmalere Skala einzuschränken, da die Personen die Items in diesem Zufriedenheitsbereich aufgrund der schlechten Differenzierung nicht genau bewerten bzw. die Natur dieser Items derartige Spielräume nicht zulässt. Auch unterschiedliche Skalenkategorienbreiten sind denkbar.

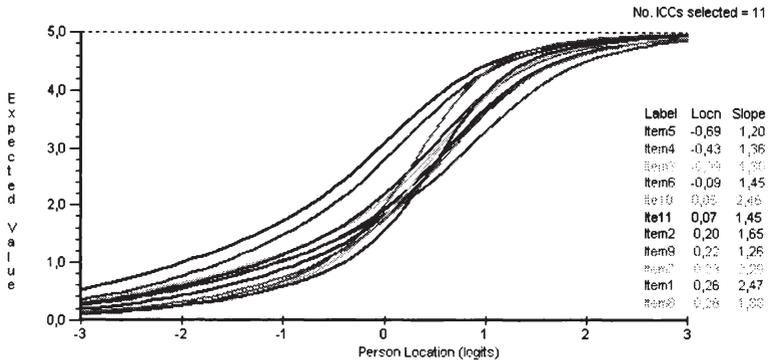


Diagramm 49: ICCs

In Diagramm 50 wird der Zusammenhang zwischen Items und Personen dargestellt. Die x-Achse stellt die Zufriedenheitsdimension dar. Einerseits befindet sich die Person auf einem gewissen Zufriedenheitslevel, andererseits werden auch die Items einem Zufriedenheitslevel zugeordnet. Dazu werden die Item-locations über die Personen-locations geplottet. Je näher die Item-locations bei den Personenlevels liegen, desto besser können die Items die Probanden untereinander differenzieren. Items mit welchen die Personen ohnehin immer zufrieden sind, vermögen dies nicht, jene mit denen niemand zufrieden ist, ebenso wenig. Eine ausgeglichene Streuung der Item-locations über die Personen-locations führt zu einer Differenzierung auf allen Zufriedenheitsstufen. Personen, welche sich auf der steilsten Stufe eines Items befinden, werden am stärksten voneinander unterschieden, da sich leichte Zufriedenheitslevelveränderungen an dieser Stelle enorm auswirken.

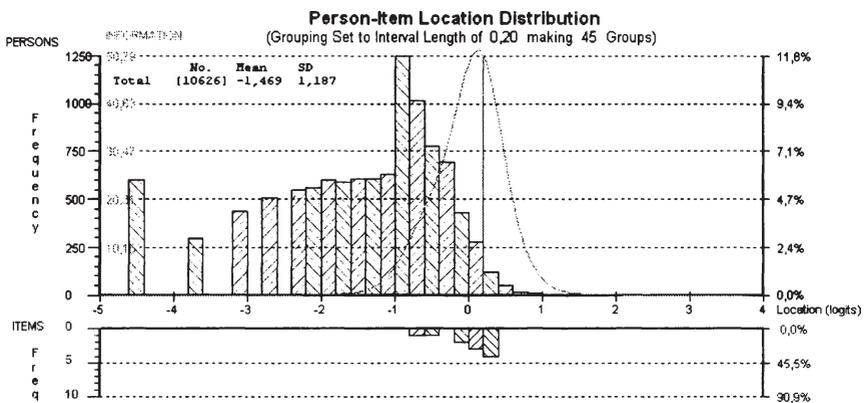


Diagramm 50: Person-Item Location Verteilung - Zufriedenheit

Die y-Achse stellt die Häufigkeiten der Probanden bzw. *locations* dar. Es zeigt sich deutlich, dass die Items im Vergleich zu den Personen sehr weit rechts angeordnet sind. Dies stellt ein klassisches Problem sozialempirischer Statistiken dar. Die Antworten auf die Items sind nicht normalverteilt sondern positiv- bzw. negativlastig. In Diagramm 51 sind die *threshold*-Levels gegen die Personenlevels geplottet. Zwar befinden sich einige der *thresholds* auch auf der linken Hälfte, nämlich jene *thresholds*, welche sich zwischen den Antwortkategorien ‚sehr zufrieden (1)‘ und 2 befinden, die ungleiche Verteilung ist jedoch deutlich erkennbar. Die *thresholds* auf der linken Seite gehören jenen Items an, für welche es äußerst schwer ist, eine sehr zufriedene Antwort zu geben. In den vergangenen Absätzen wurde dies für Buffet verdeutlicht.

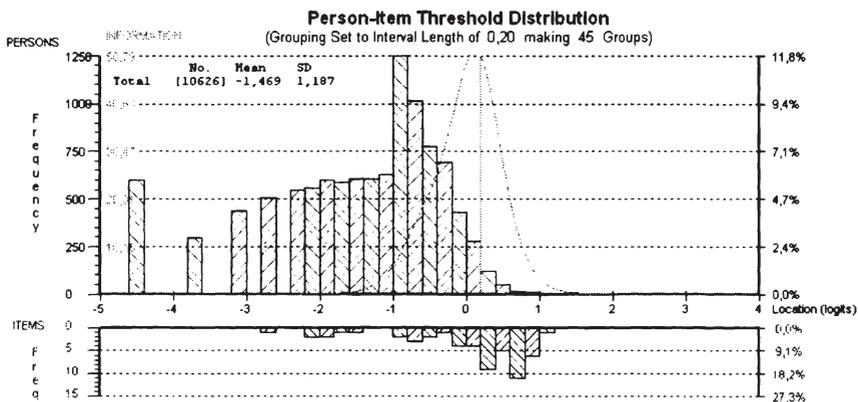


Diagramm 51: Person-Item Threshold Verteilung

Aufgrund der Itemunterschiede werden die *threshold*-Differenzen im folgenden PCM wiederum unterschiedlich geschätzt und zusammen mit den *locations* in Tabelle 39 aufgelistet. Dies bedeutet, dass die Steigung der Kurven unterschiedlich ausfallen kann.

Im Vergleich zum RSM zeigen sich beim Itemvergleich unterschiedliche *threshold*-Differenzen zwischen *threshold* eins und zwei, zwei und drei, drei und vier, und vier und fünf. Der nicht ordinale Abfall des *threshold* fünf wird in Diagramm 52 verdeutlicht.

Items	Location	Thres- hold 1	Thres- hold 2	Thres- hold 3	Thres- hold 4	Thres- hold 5
Technik/Bild/Ton	0.61388	-0.42154	1.02278	1.07208	0.76842	0.62767
Komfort	0.44394	-1.00932	0.46098	1.05635	1.06570	0.64599
Servicepersonal	0.21628	-1.50607	0.04300	1.13035	1.06605	0.34807
Preis/Leistung	-0.08020	-1.72040	-0.41454	0.58594	0.84916	0.29883
Buffet	-0.38632	-2.28264	-0.75068	0.35892	0.63340	0.10939
Atmosphäre	0.11130	-1.51394	0.00083	0.82085	0.91809	0.33068
Standort/ Erreichbarkeit	0.50477	-0.30016	0.50024	0.83881	0.84082	0.64416
Filmangebot	0.56040	-0.69138	0.45298	0.94715	1.09328	0.99995
Informations- angebot	0.42573	-1.36003	0.11111	1.13739	1.29071	0.94948
Reservierungs- möglichkeiten	0.31543	-0.45444	0.50669	0.91969	0.80524	-0.20001
Erscheinungsbild/ Image	0.24027	-1.53255	0.25074	1.01116	0.70176	0.77026

Tabelle 39: Locations und Thresholds des PCM

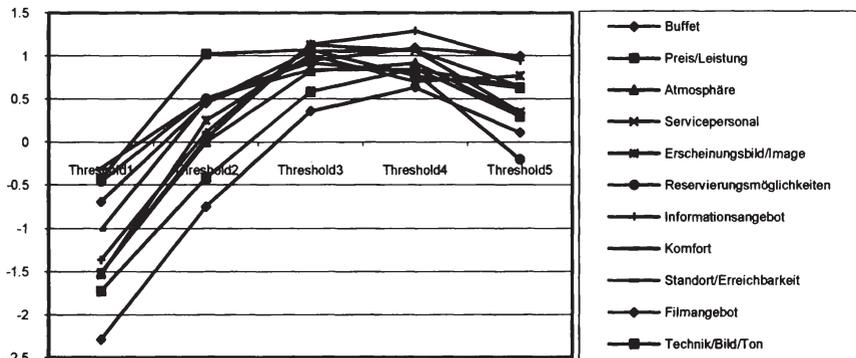


Diagramm 52: Thresholds des PCM

Die PCM-locations in Diagramm 53 geben Auskunft über die Zufriedenheitslevelanordnung der elf Items. Für Technik/Bild/Ton werden am leichtesten positive Antworten bzw. am schwierigsten negative Antworten gegeben. Für Buffet werden am leichtesten negative Antworten bzw. am schwierigsten positive Antworten gegeben. Je näher sich eine Person an den Item-locations befindet, desto stärker findet das Item eine Unterscheidung zwischen dieser Person und deren nächstgelegenen Nachbarn. Je weiter die Personen von der

Item-*location* entfernt sind, desto weniger differenziert das Item diese aus. In letzterem Falle kann sich der Zufriedenheitslevel einer Person zwar ändern, aber das Antwortresultat wird ohne enorme Veränderungen gleich bleiben.

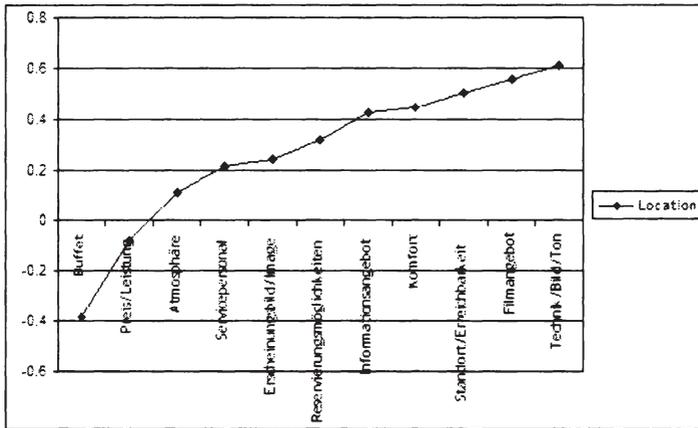
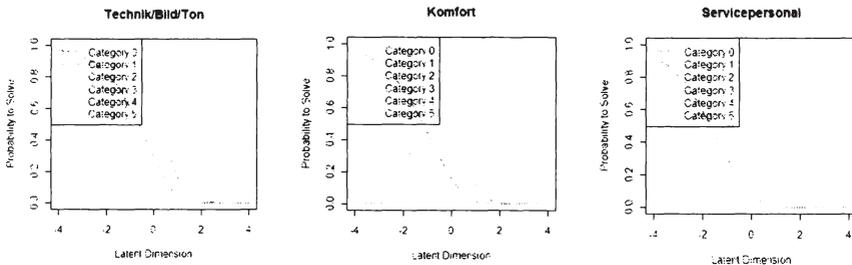


Diagramm 53: Locations des PCM

Diagramm 54 zeigt die ICCs der elf Zufriedenheitsitems. Es ist erkennbar, dass für viele Items die Wahrscheinlichkeit, Antwortkategorien im Unzufriedenheitsbereich zu wählen, niemals am größten ist. Manche der Items umfassen nur drei Kategorien, z.B. Standort/Erreichbarkeit. Manchen gelingt es besser ein breiteres Spektrum der Zufriedenheitslevels der Probanden abzubilden. Dies wird durch die Spannweite zwischen Category 0 und Category 5 auf der x-Achse dargestellt. Je näher diese beisammen liegen, desto schmaler das Spektrum, je weiter auseinander, desto breiter. Schlussfolgernd muss nochmals angemerkt werden, dass die Zufriedenheitsmessung dieser elf Items keiner 6-Punkt-Likert-Skala bedarf, sondern eine 4- oder 3-Punkt-Likert-Skala wohl ausreichend ist.



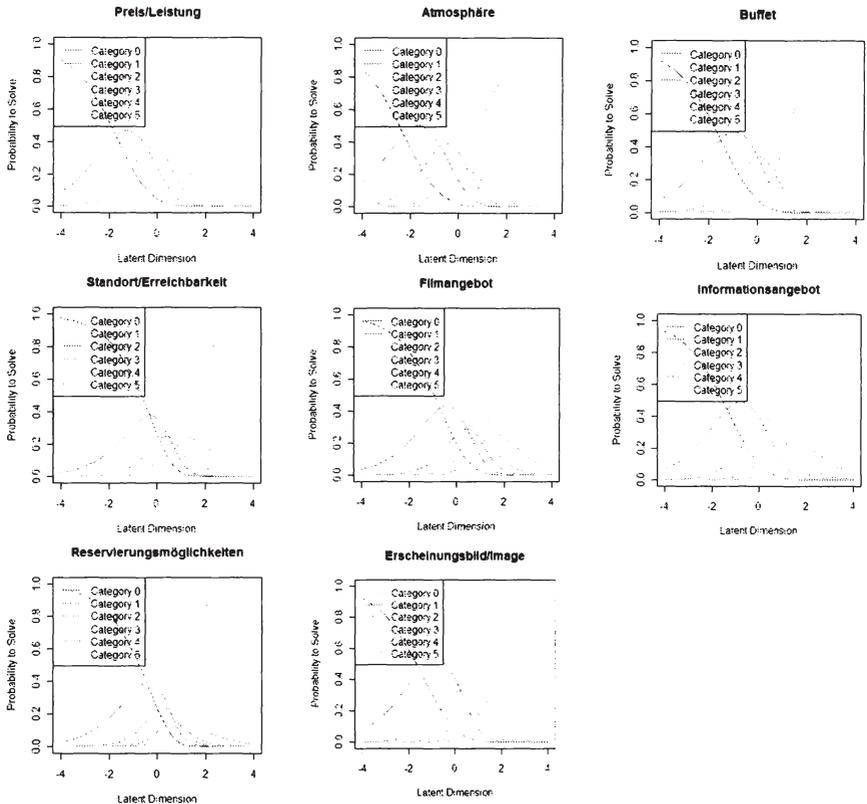


Diagramm 54: ICC Plots des PCM

Die Unterschiede und Probleme in den Kategorienantworten können auch aus anderen externen Gründen wie männlich/weiblich, loyal/nicht-loyal, etc. herrühren. Hier wird Standort/Erreichbarkeit aufgegriffen und in Tabelle 40 *locations* und *thresholds* für zwei unterschiedliche Standorte ausgegeben bzw. in Diagramm 55 visualisiert. Kinostandort 3 liegt im Zentrum der Stadt Wien, Kinostandort 18 am Rande bzw. außerhalb der Stadt.

Standort/Erreichbarkeit	Location	Thres- hold 1	Thres- hold 2	Thres- hold 3	Thres- hold 4	Thres- hold 5
Standort 3	1.07123	-0.19339	0.79013	1.44912	0.93073	2.37955
Standort 18	0.33443	-0.50117	0.48108	0.74991	0.91652	0.02584

Tabelle 40: Location- und Thresholdvergleich zweier Standorte für das Item Standort/Erreichbarkeit

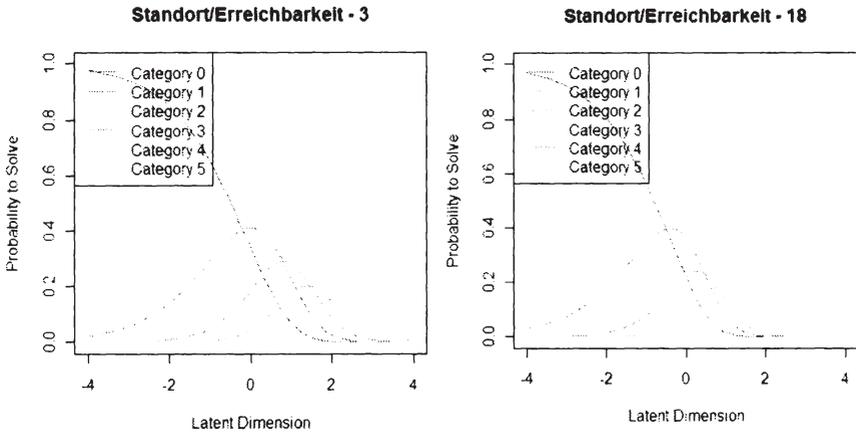
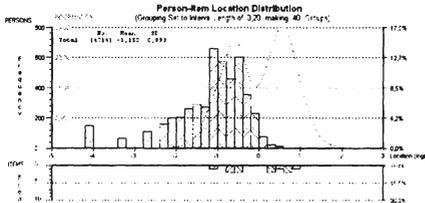


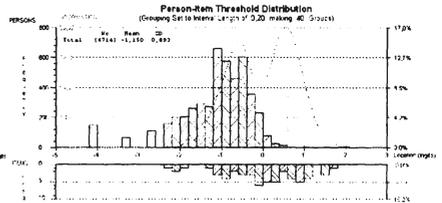
Diagramm 55: ICC-Vergleich zweier Standorte für das Item Standort/Erreichbarkeit

Standort 3 deckt ein breiteres Zufriedenheitsspektrum ab. Standort 18 beschränkt sich auf drei Antwortmöglichkeiten. Der außerhalb liegende Standort 18 weist kurz gesagt entweder positive oder negative Zufriedenheit auf. Entweder die Zufriedenheit des Probanden wird aufgrund der Lage gestört, oder die Lage ist den Probanden gleichgültig. Beim zentral gelegenen Standort 3 scheinen auch kürzere Distanzüberbrückungen eine Rolle zu spielen. Weiters befindet sich beim Standort 18 der Übergang zwischen der Kategorie ‚nicht zufrieden‘ zu einer besseren Bewertung weiter rechts. Im Vergleich zu Standort 3 wird der Zufriedenheitslevel eine positive Antwort zu geben mehr beansprucht, was auch durch die *location* zum Ausdruck kommt.

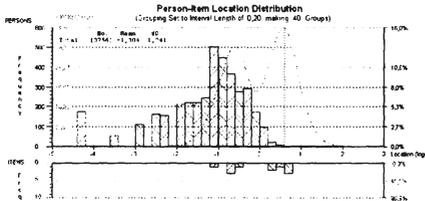
Einige der zuvor genannten Aspekte werden nun auch anhand der Wichtigkeitsbewertungen im November 2006 und September 2007 für die elf Serviceleistungsisems veranschaulicht und liefern ähnliche Erkenntnisse. Es werden wiederum die Personen-*locations* gegen die Item-*locations* und Item-*thresholds* in Diagramm 56 geplottet. Es zeigt sich, dass auch diese Items sehr positivlastig angeordnet sind. Nur wenige der Wichtigkeitsitems befinden sich in Bereichen, in welchen es diese auch vermögen, die befragten Personen untereinander zu differenzieren. Somit wird zuvor genanntes sozialstatistisches Problem zusätzlich zur Zufriedenheits- auch für die Wichtigkeitsitembatterie gezeigt. Es finden sich auch zeitliche Unterschiede in Bezug auf *locations* und *thresholds*, welche später anhand einer Modellerweiterung genauer untersucht werden.



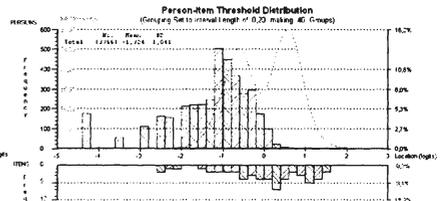
November 2006



November 2006



September 2007



September 2007

Diagramm 56: Person-Item Location Verteilung - Wichtigkeit (November 2006)

Die Wichtigkeits-*thresholds* und *locations* in Diagramm 57 geben zusätzlich Auskunft darüber, welche der Items bzw. *thresholds* in weniger wichtigen Bereichen differenzieren und welche konstant als wichtig erachtet werden. Items mit einem niedrigen *location*-Wert werden leichter als unwichtig betrachtet, als jene mit einem hohen. Buffet weist den geringsten *location*-Wert auf und wird eher als unwichtig erachtet. Auch der zugehörige tiefste *threshold* liegt im untersten Bereich. Wie bei der Zufriedenheitsmessung kann auch hier von keiner konstant ordinalen Anordnung der Kategorien gesprochen werden, da die höheren *thresholds* sinkende Werte zeigen.

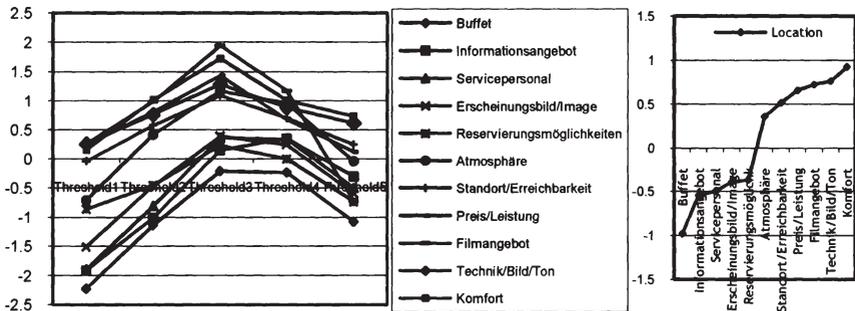


Diagramm 57: Item Thresholds und Locations – Wichtigkeiten (November 2006)

Die gewonnenen Erkenntnisse werden in Diagramm 58 für die IPA verwendet. Die *locations* aller Beobachtungen der elf Zufriedenheitsitems stehen für die Performanacewerte, die *locations* der elf Wichtigkeitsitems aus November 2006 für die Importancewerte. Die Schnittpunkte der beiden Achsen stellen jeweils die Mittelwerte der Wichtigkeits- bzw. Zufriedenheits-*locations* dar. Im rechten oberen Eck befinden sich jene Items mit welchen die Probanden zufrieden sind und welche sie auch als wichtig erachten, also Technik/Bild/Ton, Filmangebot, Komfort, Standort/Erreichbarkeit. Im linken oberen Eck befinden sich jene Items mit welchen die Probanden zwar zufrieden sind, welche sie aber nicht als wichtig erachten, also Reservierungsmöglichkeiten, Informationsangebot. Im rechten unteren Eck befinden sich jene Items, mit welchen die Probanden nicht zufrieden sind, welche sie aber als wichtig erachten, also Atmosphäre, Preis/Leistung. Im linken unteren Eck befinden sich jene Items, mit welchen die Probanden nicht zufrieden sind, welche sie aber auch nicht als wichtig erachten, also Servicepersonal, Erscheinungsbild/Image, Buffet. Diese Einteilung kann analog zur originalen IPA von Martilla und James aus dem Jahre 1977 dafür verwendet werden, um Items zu identifizieren, bei welchen Verbesserungsbedarf besteht bzw. Erwartungen übererfüllt werden.

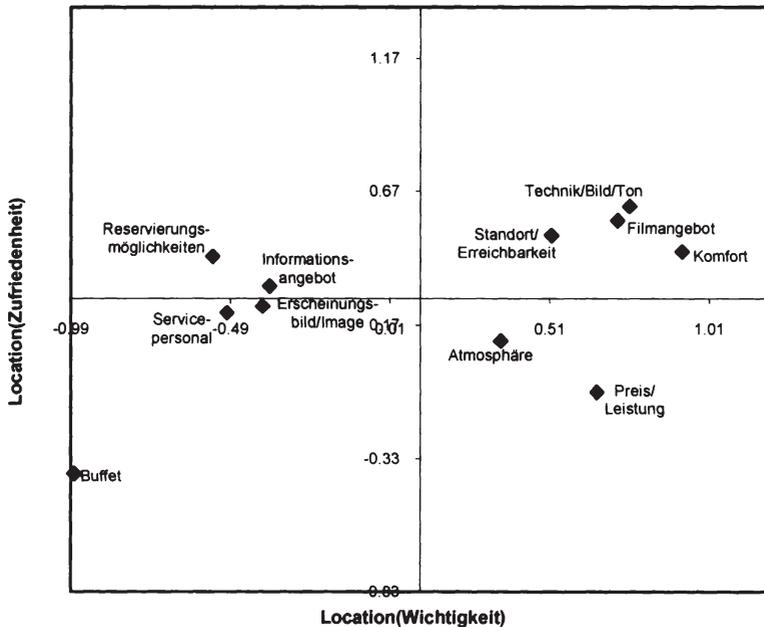
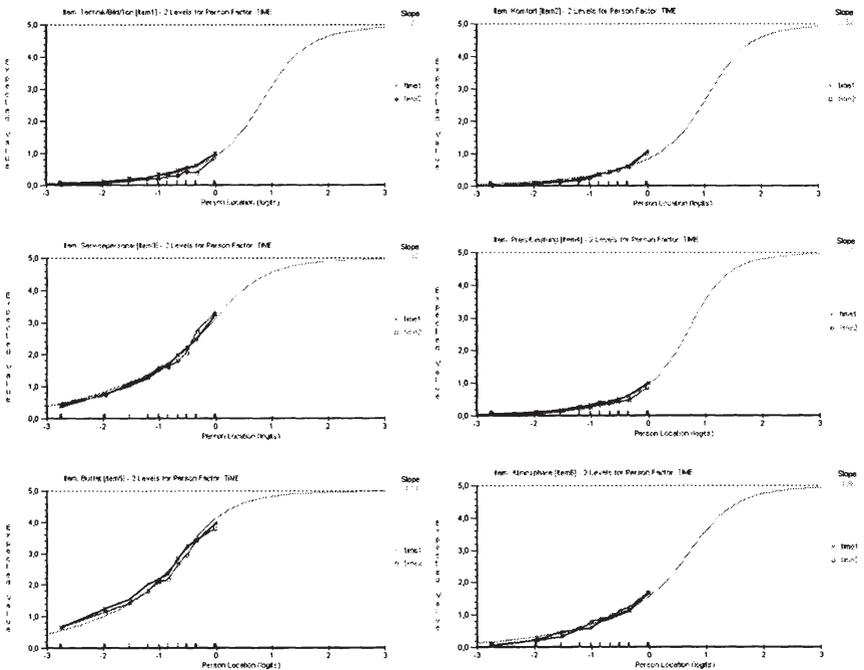


Diagramm 58: IPA mit Zufriedenheits- und Wichtigkeits-Locations

4.3.2.5. Gruppenvergleiche - Kinodaten

Über einen Gruppenvergleich wird hier ein Übergang zu zeitlichen Veränderungen geschaffen. Gruppe eins beinhaltet die Wichtigkeitsbewertungen aus Oktober 2006 und Gruppe zwei aus September 2007. Lediglich in dieser Analyse werden die Personen zu den beiden Zeitpunkten als nicht ident betrachtet. Diese Informationsverkürzung wird später beseitigt. Diagramm 59 zeigt die Kurven der Erwartungswerte und Personen-locations beider Gruppen. Zum Zeitpunkt zwei werden für Technik/Bild/Ton leichter positive Wichtigkeitswerte angegeben. Komfort weist keine gruppenzeitlichen Unterschiede auf. Die Reservierungsmöglichkeiten erhalten zum Zeitpunkt eins leichter positive Wichtigkeitswerte. Liegt die Kurve zum Zeitpunkt zwei tiefer, steigt die Wichtigkeit, liegt sie zum Zeitpunkt eins tiefer, sinkt diese über die zehn Monate. Wichtigkeitssinkende Abweichungen werden für Technik/ Bild/Ton, Preis/Leistung, Buffet und Informationsangebot gefunden. Wichtigkeitssteigende Abweichungen zeigen Atmosphäre, Standort/ Erreichbarkeit und Reservierungsmöglichkeiten. Die Wichtigkeiten der restlichen Items bleiben relativ gleich bzw. schwanken, je nachdem wo die Personen-locations liegen, sowohl positiv als auch negativ um deren zeitunterschiedliche Bewertungen.



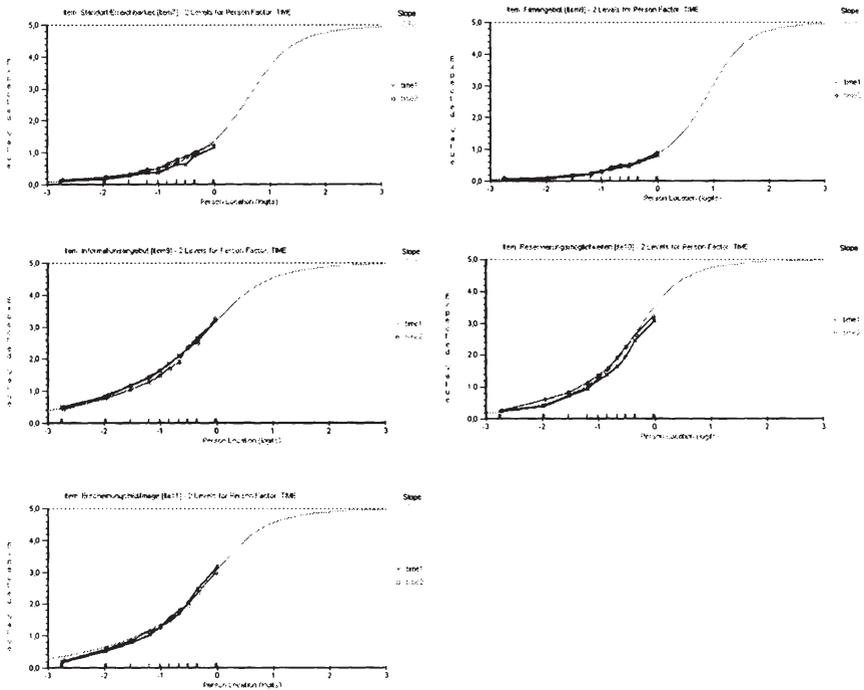


Diagramm 59: Zeitliche Gruppenlösung – Wichtigkeiten

4.3.2.6. Linear Logistic model with Relaxed Assumptions - Kinodaten

In Diagramm 60 werden mit Hilfe einer Erweiterung des dichotomen Modells, nämlich dem LLRA, personenspezifische Veränderungen über die Zeit hinweg geschätzt und visualisiert. Als Datenbasis dienen die Zufriedenheitsbeobachtungen der elf Zufriedenheitsvariablen aller zwölf Wellen. Es wird die Veränderung der Zufriedenheitsbewertung für die einzelnen Personen gemessen. Die Originaldaten werden zuvor dichotomisiert. Bewertungen eins, zwei und drei erhalten die Werte null, vier fünf und sechs die Werte eins. Es wird ohne Fehlerwerte, also mit listenweisem Fallausschluss, geschätzt. Um nicht zu viele Probanden zu verlieren wird maximal über sechs Monate geschätzt. Obere Graphik enthält 190 Personen, welche in jedem der ersten sechs Monate, Oktober 2006 bis März 2007, Zufriedenheitsbewertungen abgegeben haben. Mittlere Graphik enthält 135 Probanden mit vollständigen Bewertungen in den ungeraden Wellen der Monate Oktober und Dezember 2006, Februar, April, Juni und August 2007. Untere Graphik enthält 171 Probanden aus den geraden

Wellen der Monate September 2006, November 2006, Jänner, März, Mai, Juli und September 2007.

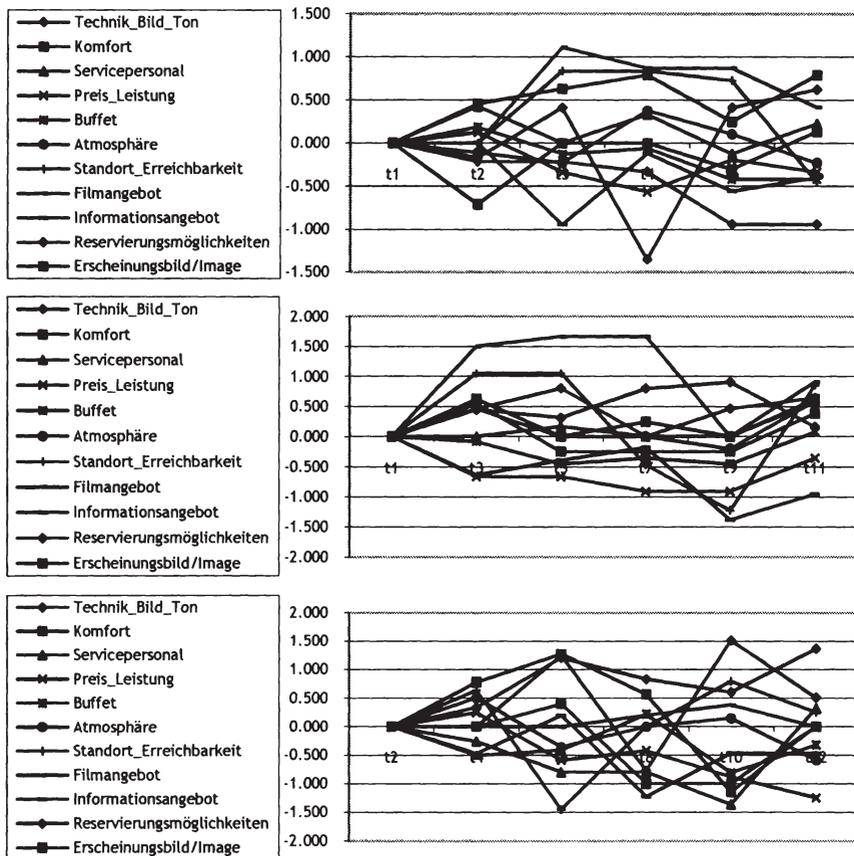


Diagramm 60: LLRA - Zufriedenheit

Über die ersten sechs Wellen zeigen sich z.B. negative Bewertungstrends für Preis/Leistung und Reservierungsmöglichkeiten. Es wird schwerer, diese Probanden zufriedenzustellen. Für Filmangebot und Erscheinungsbild/Image finden sich hingegen positivere Werte über den Zeitraum und es wird leichter die Probanden zufriedenzustellen. Über die ungeraden Wellen zeigen sich für Informationsangebot und Preis/Leistung negative Trends und für Technik/Bild/Ton ein positiver Trend. Über die geraden Wellen zeigen sich negative Trends für Buffet, Preis/Leistung und Filmangebot und ein positiver Trend für die Reservierungsmöglichkeiten.

Nun wird vorangegangenes LLRA Modell anhand der Wichtigkeitsbewertungen durchgeführt. Die Frage ist, ob sich zusätzlich zu den Zufriedenheitsbewertungen auch bei den Wichtigkeitsbewertungen Veränderungen über die Zeit aufspüren lassen. Vor der Visualisierung der Veränderungen wird in Tabelle 41 anhand von Kontingenztafeln dargestellt, wie sich die Anzahl an positiven, Werte mit 0, und negativen Bewertungen, Werte mit 1, über die Zeit hinweg verändern.

Technik/Bild/Ton				Komfort				Servicepersonal			
		Zeitpunkt 1				Zeitpunkt 1				Zeitpunkt 1	
		0	1			0	1			0	1
Zeit-	0	3112	22	Zeit-	0	3144	21	Zeit-	0	2324	254
punkt 2	1	48	8	punkt 2	1	20	5	punkt 2	1	343	269
Preis/Leistung				Buffet				Atmosphäre			
		Zeitpunkt 1				Zeitpunkt 1				Zeitpunkt 1	
		0	1			0	1			0	1
Zeit-	0	3122	23	Zeit-	0	1679	259	Zeit-	0	3014	65
punkt 2	1	41	4	punkt 2	1	530	722	punkt 2	1	88	23
Standort/Erreichbarkeit				Filmangebot				Informationsangebot			
		Zeitpunkt 1				Zeitpunkt 1				Zeitpunkt 1	
		0	1			0	1			0	1
Zeit-	0	3071	46	Zeit-	0	3138	25	Zeit-	0	2200	271
punkt 2	1	63	10	punkt 2	1	24	3	punkt 2	1	443	276
Reservierungsmöglichkeiten				Erscheinungsbild/Image							
		Zeitpunkt 1				Zeitpunkt 1					
		0	1			0	1				
Zeit-	0	2306	269	Zeit-	0	2384	270				
punkt 2	1	307	308	punkt 2	1	333	203				

Tabelle 41: Kontingenztafel - Wichtigkeit

In den Zeilen stehen die positiven und negativen Bewertungen aus November 2006, in den Spalten die aus September 2007. Der Wert in jener Zelle, wo sich Spalte null und Zeile null überschneiden, gibt die Anzahl der Bewertungen aus, in welchen zu beiden Zeitpunkten positive Antworten gegeben wurden. Der Wert in jener Zelle, wo sich Spalte eins und Zeile eins überschneiden, gibt die Anzahl der Bewertungen aus, in welchen zu beiden Zeitpunkten negative Antworten gegeben wurden. Der Wert in jener Zelle, wo sich Spalte null und Zeile eins überschneiden, gibt die Anzahl der Bewertungen aus, in welchen zum Zeitpunkt zwei negative Antworten und zum Zeitpunkt eins positive Antworten gegeben

wurden. Der Wert in jener Zelle, wo sich Spalte eins und Zeile null überschneiden, gibt die Anzahl der Bewertungen aus, in welchen zum Zeitpunkt zwei positive Antworten und zum Zeitpunkt eins negative Antworten gegeben wurden. Befinden sich im Tabellenfeld links oben weniger Werte, verglichen mit Tabellenfeld rechts unten, deutet dies darauf hin, dass zum zweiten Zeitpunkt mehr Probanden das Item als unwichtig erachtet haben, im Vergleich zum ersten Zeitpunkt, wo dies noch als wichtig erachtet wurde.

Auffällig ist, dass es bei weitem weniger negative Bewertungen gibt, als positive. Dies ist deutlich zu beiden Zeitpunkten erkennbar. Auch zwischen den Items sind Unterschiede auffindbar. In den Items Servicepersonal, Buffet, Informationsangebot, Reservierungsmöglichkeiten und Erscheinungsbild/Image tauchen weit mehr negative Bewertungen auf, verglichen mit den restlichen Items. Trotzdem wird keine ausgeglichene Bewertung positiver und negativer Antworten erreicht. Diagramm 61 zeigt die Ergebnisse des LLRA.

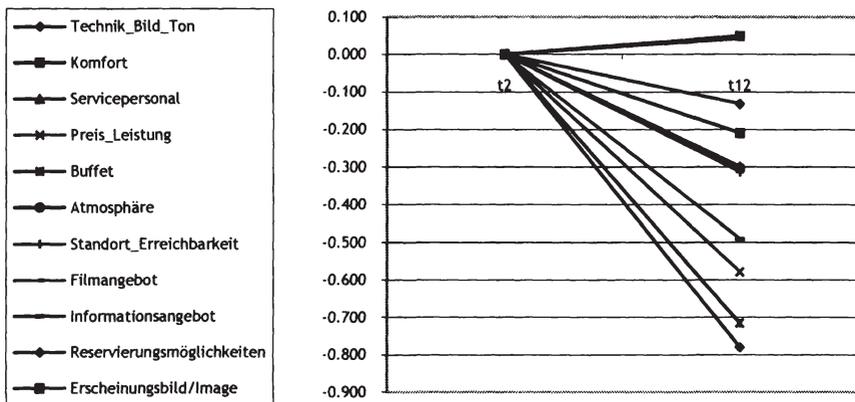


Diagramm 61: LLRT-Wichtigkeit

Hierbei muss beachtet werden, dass alle Veränderungsparameter als eigene Dimension geschätzt werden. Somit wirken auf alle elf Items unterschiedliche Kräfte. Es stellt sich die Frage, ob ein allgemeiner Trend, also eine Dimension, für alle elf Items ausreichend ist. Auch für einzelne Itemgruppen könnten nämlich Trendparameter berechnet werden. Die Parameter von Servicepersonal, Atmosphäre und Informationsangebot verhalten sich sehr ähnlich und für diese drei Items kann möglicherweise eine eigene Dimension im Gesamtmodell aufgenommen werden, die die drei einzelnen Dimensionen ersetzt und das Modell, aufgrund der Einsparung von Parametern, verbessern. Auch Komfort

und Filmangebot verhalten sich ähnlich. Hier wird als Vergleichsmodell eine Dimension, also ein einheitlicher Trendparameter für alle Items, geschätzt, um eine Aussage darüber zu treffen, ob eine Differenzierung zwischen den Items sinnvoll ist. Das Ergebnis liefert einen Wert von -0.38448197 bei einer Standardabweichung von 0.03319857. Es zeigt sich ein allgemeiner negativer Trend. Folglich wird untersucht, ob dieser Trend signifikant auf alle elf Items angewandt werden kann. Dazu wird das erste Modell mit den elf unterschiedlichen Dimensionen, mit dem zweiten Modell mit nur einer Dimension, verglichen. Die *likelihood ratio statistic* ist mit einem Wert von 43.85843 bei 10 Freiheitsgraden signifikant, $p = 0$. Es besteht folglich ein signifikanter Unterschied zwischen beiden Modellen. Der Trendparameter kann nicht für alle Items vereinheitlicht werden und eine Ausdifferenzierung zwischen den Items erscheint sinnvoll.

4.3.3. Multiple Korrespondenzanalyse

Ziele:

Die den Teilzufriedenheitsvariablen zugrundeliegenden Dimensionen werden unter Berücksichtigung der Skalierung ermittelt und verwendet um die Gesamtzufriedenheit vorherzusagen. Den Daten innewohnende Probleme werden aufgedeckt und sollten bei der Interpretation nachfolgender Analyseergebnisse berücksichtigt werden.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...?
(gemessen im Monat September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image, Film;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Bitte machen Sie folgende Angaben zu Ihrem Kinobesuch: Gesamtzufriedenheit
(gemessen im Monat September 2007);

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Programm:

R (Package: homals);

Modell:

Die im Folgenden vorgestellten Modelle zählen zum Bereich des *optimal scaling* und werden bereits umfangreich in einem Werk von Gifi (1981) beschrieben.

Mit Hilfe einer *loss*-Funktion und einem *alternating least squares*-Verfahren (ALS) wird die Transformation der Variablen in einen numerischen Raum vorgenommen. Für jede Dimension werden optimal skalierte Kategorien- und Objektwerte berechnet. Der mehrdimensionale Raum wird folglich auf einem Raum mit weniger Dimensionen abgebildet. Es werden unterschiedliche Einschränkungen vorgenommen. Beispielsweise werden über Rangrestriktionen nicht-lineare Hauptkomponentenanalysen durchgeführt. Zum Zwecke einer nicht-linearen kanonischen Korrelationsanalyse werden Restriktionen für einzelne Variablengruppen eingeführt. Durch die Gruppierung von Variablen in Sets werden nicht-lineare kanonische Korrelationen bzw. Modelle zur Prognose wie in der Regressionsanalyse oder der Diskriminanzanalyse geschaffen. Solange die Likert-Skalen des vorliegenden Kinodatensatzes als ordinal behandelt werden, gibt es aus statistischer Sicht, betreffend der Sauberkeit der Analysen, auch keine großen Probleme. Oft werden sie jedoch zwangsbedingt aufgrund der Schätzverfahren unterschiedlicher Methoden als intervallskaliert betrachtet. Aufgrund der meist unterschiedlichen Abstände zwischen den Kategorien einer Ordinalskala kann dies jedoch schwer angenommen werden. Um die ordinal skalierten Variablen in einen intervallskalierten Raum optimal zu reskalieren, können die Abstände zwischen den Kategorien auseinandergezogen bzw. zusammengestaucht werden. Ein Vergleich zwischen der Behandlung der Daten als ordinal, verglichen mit der Annahme, die Daten wären numerisch skaliert, wird im Zuge der folgenden Analysen vorgestellt. Die gezeigten Modelle können für kategorial geordnete als auch kategorial nicht geordnete (nominale) Skalenwerte verwendet werden. In unterschiedlichen Programmen (R, SPSS, etc.) kann die Homogenitätsanalyse bzw. multiple Korrespondenzanalyse unter dem Begriff *homals* aufgefunden werden. Anschauliche Beispiele mit Erklärungen zum Modell finden sich in deLeeuw und Mair (2008), Mair und deLeeuw (2008) oder auch Michailidis und deLeeuw (1998). Im Zuge der Homogenitätsanalyse werden einerseits die Korrelationen zwischen den Items eines multivariaten Datensatzes maximiert, gleichzeitig werden optimale Skalenwerte erzeugt. Die Korrespondenzanalyse wird normalerweise zur deskriptiven Beschreibung kategorialer Daten verwendet. Beim hier verwendeten multivariaten Datensatz dient die Homogenitätsanalyse dem Hauptziel der Homogenitätsoptimierung.

Werden kontinuierliche Variablen als kategorial betrachtet, stellt jeder metrische Wert eine Kategorie dar. Die Verbindungen zwischen Objekten, hier Probanden, und den Kategorien der Items können, um Zusammenhänge zwischen diesen zu erkennen, lediglich über eine Kontingenztafel visualisiert werden. Mit steigender Anzahl an Objekten und Kategorien wird diese Tabelle unüber-

sichtlich. Deshalb wird hier der Ansatz eines euklidischen Dimensionsraumes verfolgt, wo sowohl Objekte als auch Kategorien eingezeichnet werden. Der logischen Abfolge nach werden zuerst alle Objekte mit den jeweils gewählten Kategorien verbunden. Je kürzer die Verbindungen, desto besser, da die Objekte nahe ihrer zugehörigen Kategorien positioniert werden sollen. Mit Hilfe der *loss*-Funktion werden eben diese Abstände zwischen Objekten und zugehörigen Kategorien minimiert, was später im *graphplot* veranschaulicht wird.

Korrespondenzanalysemodelle werden normalerweise über die Singulärwertzerlegung gelöst. Diese berechnet eine orthonormale Basis für den Spaltenraum einer Matrix, der linke Singulärvektor. Das *homals*-Paket in R löst das Problem mit Hilfe der *ALS-loss*-Funktion. Dabei wird die Abweichung von der Homogenität minimiert. Ein mehrdimensionaler Raum wird aus 0-1 codierten Daten auf einen niedriger dimensionalen Raum mit orthogonal angeordneten Spalten projiziert. Die 0-1 codierte Datenmatrix gibt an, ob für die jeweilige Variable die jeweilige Kategorie gewählt wurde. Dabei werden eine $n \times p$ -Matrix für die Objektwerte und eine $k_j \times p$ -Matrix für die Kategorienwerte erzeugt. n bestimmt die Anzahl an befragten Personen, k die sechs Kategorien der Likert Skala, p die Anzahl an Dimensionen, und $j = 1, \dots, m$ steht für die unterschiedlichen Zufriedenheitsvariablen. Zuerst wird nach den Kategorienwerten der einzelnen Variablen mit fixierten Objektwerten minimiert. Anschließend wird nach den Objektwerten mit fixierten Kategorienwerten minimiert. Alternierend werden abwechselnd die Werte für die Kategorien und anschließend die Werte für die Objekte berechnet. Im Anschluss werden die Objektwerte, diese stellen orthogonale Spalten der Objekt- bzw. hier Personenwerte dar, zentriert und orthonormalisiert. Die Normalisierung gewährleistet die Zentrierung um den Ursprung. Dies wird so lange fortgeführt, bis das Modell bei einem zuvor festgelegten Wert konvergiert. Nach jeder Iteration wird nämlich der Wert der *loss*-Funktion berechnet, um diesen auf Konvergenz hin zu überprüfen. Sinkt der Wert nicht mehr signifikant, wird also der Wert einer vordefinierten Grenze durch den sinkenden Wert nicht mehr überschritten, stoppt der Iterationsprozess. Die Minimierung der *loss*-Funktion führt zu $n \times p$ Komponentenwerten und $m \times p$ Komponentenladungen. Die Kategorienwerte befinden sich letztendlich im Zentroid der zugehörigen Objektwerte. Die Objektwerte beschreiben den Durchschnitt der zugehörigen Kategorienwerte. Haben Personen die gleichen Antworten gegeben, so erhalten sie auch die gleichen Objektwerte. Die Eigenwerte stellen das Verhältnis der durchschnittlichen Varianz zwischen den Kategorien und der Gesamtvarianz jeder Dimension dar.

Sind die Variablen nicht-metrisch, kann eine nicht-lineare Hauptkomponentenanalyse durchgeführt werden. Dies hat die nicht-lineare Transformierung der beobachteten Variablen zur Folge. Dafür wird eine Homogenitätsanalyse mit Einschränkungen in Bezug auf die Kategorienbewertungen vorgenommen. Diese kann mit Hilfe der Gewichtsmatrix durch Beschränkung der Bewertungskategorien zwischen den Dimensionen vorgenommen werden. Der bedeutendste Spezialfall ist die sogenannte Rang-eins-Beschränkung, *rank-1*. Das Modell wird, ohne die Anzahl der Dimensionen zu reduzieren, gerechnet. Es können numerische, kategoriale, als auch nominale Skalenniveaus innerhalb eines Modells verwendet werden, indem diese jeweils als Set definiert werden. Berechnet wird jedoch lediglich ein Set an Kategorienwerten. Die Kategorienwerte der zweiten Dimension stellen aufgrund der erwähnten Gewichtung eine Linearkombination der Werte der ersten Dimension dar. Somit gibt es nur ein Set an Kategorienwerten für jede Variable, Komponentenladungen dazu jedoch für alle Dimensionen. Das bedeutet, dass es eine mehrdimensionale Lösung für die Objektwerte gibt, jedoch nur eine optimale Skalierung für die Itemkategorien. Der Schätzalgorithmus sieht ähnlich aus. Zuerst werden die multiplen Kategorienwerte berechnet, dann die Komponentenladungen und schließlich die einfachen Kategorienwerte. Das Skalenniveau wird durch monotone oder lineare Regression berücksichtigt. Nach Berechnung der multiplen Kategorienwerte werden die Objektwerte geschätzt und die Spaltenzentrierung und Orthogonalisierung der Objektwerte vorgenommen. Wiederum wird so lange alterniert, bis ein Konvergenzkriterium erreicht ist. Für ordinale Daten wird eine monotone Regression und für numerische eine lineare durchgeführt. Die *a priori* Ordnung der Ordinalskalen bleibt erhalten. Sie liegen richtig angeordnet auf einer Geraden. Numerisch skalierte Items besitzen überdies gleichmäßige Intervalle auf dieser Geraden. Dies wird später graphisch veranschaulicht.

Auch Gruppen von Variablen können definiert werden. Die sogenannte nicht-lineare kanonische Korrelationsanalyse, *overals* genannt, dient der Bildung kanonischer Korrelationsanalysen und Vorhersagemodelle. In *homals* steht jede Variable für ein Set und es werden p Dimensionen, aus den 0-1 codierten Daten, von denen es für jede Variable die Kategorienspalte gibt, gebildet. Bei der kanonischen Korrelation werden Variablensets gebildet und p orthogonale Vektoren für die jeweiligen Sets errechnet. Im Vergleich zur klassischen kanonischen Korrespondenzanalyse ist die nicht-lineare kanonische Korrespondenzanalyse nicht auf zwei Sets beschränkt. Werden asymmetrische Sets gegeneinander definiert, können auch Prognose- und Regressionsmodelle gebildet werden. Die Prädiktorvariablen, welche die Originalantworten beschreiben, werden in der nicht-linearen kanonischen Korrelationsanalyse durch

einen niedriger dimensionalen Raum an Variablen dargestellt. Die erklärte Varianz der Variablen, welche durch das komprimierte Set erklärt wird, wird dafür maximiert. Somit werden Abhängigkeiten zwischen zwei Variablensets erklärt, nach dem deren interne Abhängigkeiten beseitigt wurden. Jede Variable bringt jenen Anteil unabhängiger Information ein, welcher in Bezug zu den anderen Variablensets und auch innerhalb des eigenen unabhängig ist. Die Information wird also aus den Abhängigkeiten zwischen den Sets ermittelt. Die Korrelationen zwischen den Sets werden dafür maximiert. Die resultierende kanonische Korrelation wird, kurz gesagt, aus einer Linearkombination eines Variablensets, den kanonischen Variablen, und dem eines anderen Sets errechnet. Laut dem Zentroidprinzip sind die Kategorienwerte die Zentroide der Objektwerte, welche um den Einfluss der anderen Variablen im Set korrigiert werden. So kann auch ein weiteres Paar an kanonischen Variablen orthogonal zum ersten erzeugt werden, dies wäre die zweitgrößte Korrelation. Insgesamt wird dies solange fortgeführt, bis eine zufriedenstellende Anzahl an Dimensionen erzeugt wurde.

Weitere Restriktionen für Likert-Skalen ermöglichen eine vordefinierte Ordnung der Skalenwerte, welche durch monotone Transformation in das Modell aufgenommen wird. Ist die Skala numerisch, können die Abstände zwischen den Kategorien fixiert werden. Würde man letzte Einschränkung für eine Likert-Skala vornehmen, führt dies zu einer affinen Transformation. Die Distanzen zwischen den ursprünglichen Kategorien würden auf der transformierten Skala gleiche Werte erhalten, was normalerweise nicht angenommen werden kann. Auch Variablengruppen mit unterschiedlichen Skalen, wie nominal, ordinal, intervall oder polynomial können berücksichtigt werden. Hier werden ordinal-skalierte Variablen eingebracht, nämlich zwölf Zufriedenheitsitems als Gesamtset und die Gesamtzufriedenheit als Einzelset.

deLeeuw und Mair (2008) vergleichen die Homogenitätsanalyse beispielsweise mit der Korrespondenzanalyse. Versucht wird, aus den in Kreuztabellen enthaltenen Systemen der Reihen- und Zeilenwerte Korrelationen zwischen den Variablen zu maximieren. Die beiden Methoden sind gleich, wenn die Ähnlichkeiten über die Zeilen gerechnet werden. Die Korrespondenzanalyse benötigt symmetrische Datenmatrizen, die Homogenitätsanalyse kann unterschiedliche Einträge enthalten. Ein weiterer Vergleich ist über die multidimensionale Skalierung möglich. Diese passt ebenfalls Unähnlichkeiten mehrerer Objekte in einem Raum mit weniger Dimensionen an. Die Zeilen werden dabei als Profile verwendet. Verglichen mit der Clusteranalyse, wo es darum geht, welche Kategorien von wie vielen und welchen Personen gewählt

Diagramm 63 bildet die Teilzufriedenheitsvariablenkategorien in der jeweiligen Graustufe ab. Im oberen Teil befinden sich die schlechten Zufriedenheitsbewertungskategorien. Je näher man sich an den unteren Rand der dichten Wolke begibt, desto wahrscheinlicher wird es, eine sehr zufriedene Bewertung anzutreffen. Die weiteren Nummerierungen stehen wiederum für die Personen. Auch die Gesamtzufriedenheit, *overall*, wird dargestellt, fließt jedoch nicht in die Modellberechnung ein.

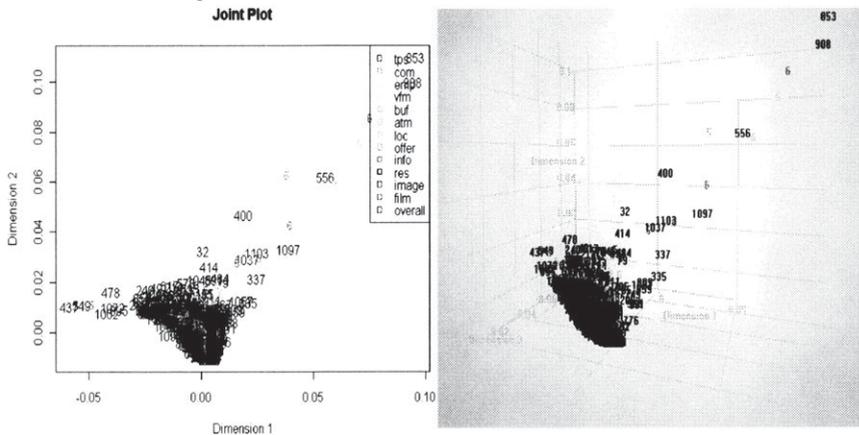


Diagramm 63: Joint Plot

Diagramm 64 zeigt *span plots* der Dimensionspaare für das Item Film, in welchen jeder Punkt eine Person darstellt. Je nachdem welche Antwort die Person gegeben hat, wird die Verbindung zum nächstgelegenen Probanden der gleichen Antwortkategorie, mit einer Linie verbunden. In allen drei Plots ist erkennbar, dass sich die Räume des Plots, welche die unterschiedlichen Kategorien besetzen, nicht voneinander unterscheiden. Die Kategorienräume überschneiden sich und bilden lediglich eine einzige Punktwolke. Die drei Dimensionen vermögen es nicht, die Personen anhand der Kategorien der Zufriedenheitsvariable Film zu unterscheiden.

Diagramm 65 zeigt die Ladungsrichtungen der drei Dimensionen auf die Teilzufriedenheitsvariablen. Die drei Dimensionen beschreiben inhaltlich jeweils ähnliche Aspekte. Sie grenzen die Ladungen voneinander ab, welche vor, und solche die im Zuge des Kinobesuches von Relevanz sind.

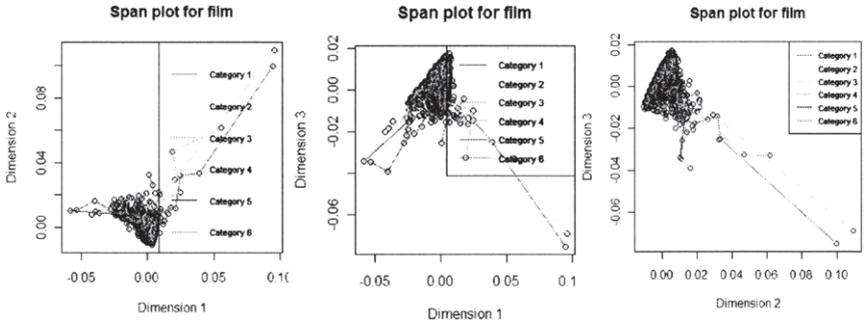


Diagramm 64: Span Plot

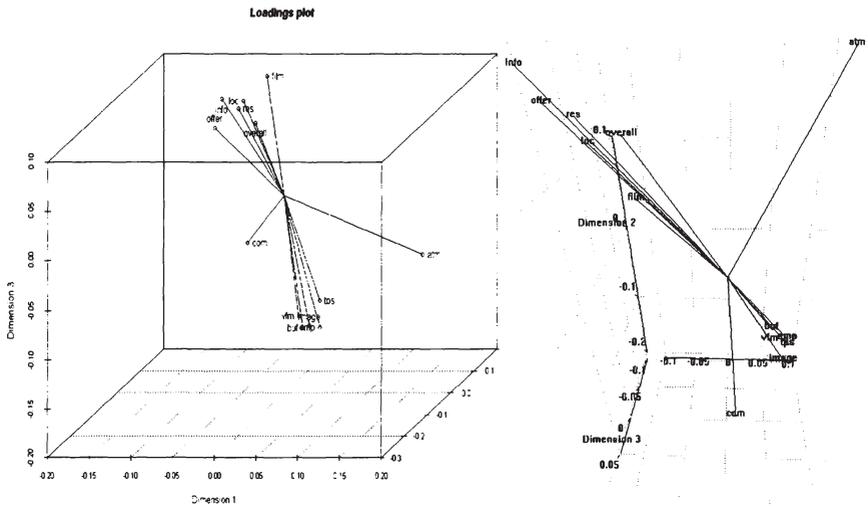


Diagramm 65: Ladungsplot

Aus der soeben visualisierten Lösung werden die Klassifizierungsprozentwerte in Tabelle 42 errechnet, wobei die Personen ihrer jeweils nächstgelegenen Kategorie für die einzelnen Items zugeordnet werden. Am besten vermögen es die drei Dimensionen die Personen nach deren Kategorienzugehörigkeit für das Item Erscheinungsbild/Image zu trennen.

In Diagramm 66 werden sämtliche Personen mit deren jeweils original gewählten Variablenkategorie verbunden. Die nummerierten Sterne stellen Personen dar. Die Punkte stehen für die Kategorien der zwölf Items. Die Zuordnung über die schwarzen Verbindungslinien stellt die Originalzuordnung zu den Kategorien dar, verglichen mit vorheriger Tabelle, wo die Zuordnung

über die nächstgelegene Kategorie erfolgte. Es entsteht eine dichte Wolke. Je ähnlicher sich die Personen sind, desto weniger Linien benötigt dieser Plot, da sich die Linien dann überschneiden. Die Linien werden auch kürzer, wenn die Personen von den zugehörigen Kategorien weiter entfernt sind. Je weniger Farbe also benötigt wird, desto besser die Zuordnung. Sind die Kategorienwerte eines Items weit voneinander entfernt, so diskriminiert diese Variable die Objekte besser.

Variable	Klassifizierungsrate	Variable	Klassifizierungsrate
Technik/Bild/Ton	64,54 %	Standort/Erreichbarkeit	57,21 %
Komfort	64,90 %	Filmangebot	61,01 %
Servicepersonal	61,27 %	Informationsangebot	67,37 %
Preis/Leistung	54,11 %	Reservierungsmöglichkeiten	58,18 %
Buffet	51,90 %	Erscheinungsbild/Image	74,45 %
Atmosphäre	68,08 %	Film	56,06 %
Gesamtzufriedenheit	52,52 %		

Tabelle 42: Klassifizierungsrate

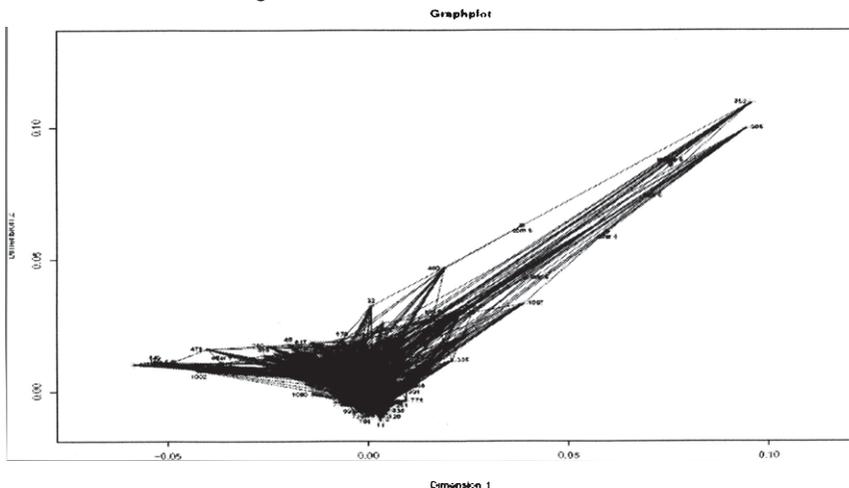


Diagramm 66: Graphplot

In Diagramm 67 wird das Item Filmangebot visualisiert. Die gewählten Kategorien werden mit den Objektstandorten verbunden. Je weniger sich die Linien überschneiden, desto trennschärfer sind die Kategorien der Variablen, da unterschiedliche Antwortkategorien dann nicht mehr nahe beieinander liegen. Es

muss beachtet werden, dass die Zuordnung und Lageberechnung der Personen die Qualität des Gesamtmodells widerspiegelt. Wiederum ist ersichtlich, dass sich die zufriedeneren Kategorien sehr stark überschneiden.

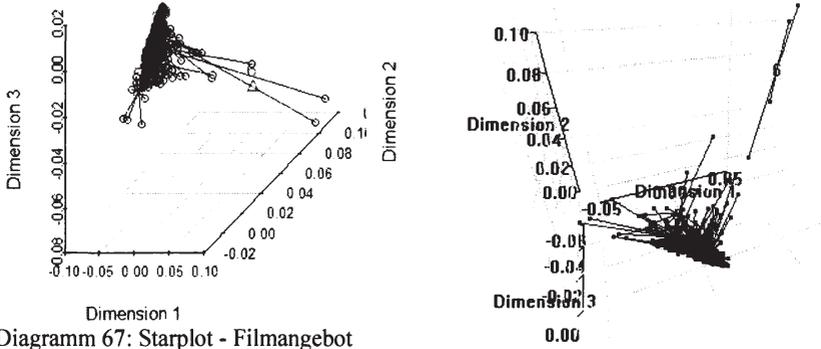


Diagramm 67: Starplot - Filmangebot

Im Folgenden wird versucht die Gesamtzufriedenheit über die zwölf Zufriedenheitsitems zu erklären. Dazu werden die zwölf Items als Set definiert und die Gesamtzufriedenheit alleine als eigenes Set. Tabelle 43 zeigt anhand der Prognosefunktion, welche Zufriedenheitsitems durch das Modell am besten abgebildet werden. Dies ist aufgrund der Bestimmung der Gesamtzufriedenheitsvariablen als eigenes Set, die Gesamtzufriedenheit. Von den Teilzufriedenheiten ist das Item Film mit 63,4 % die am besten repräsentierte.

Variable	Klassifizierungsrate	Variable	Klassifizierungsrate
Technik/Bild/Ton	58.80 %	Standort/Erreichbarkeit	41.64 %
Komfort	47.83 %	Filmangebot	36.07 %
Servicepersonal	44.39 %	Informationsangebot	43.59 %
Preis/Leistung	35.10 %	Reservierungsmöglichkeiten	41.64 %
Buffet	31.12 %	Erscheinungsbild/Image	45.98 %
Atmosphäre	48.54 %	Film	63.40 %
Gesamtzufriedenheit	82.49 %		

Tabelle 43: Prognosetabelle

Diagramm 68 gibt die Lage der Gesamtzufriedenheitsgruppenkategorien wider. In vorangegangem Voronoi Plot werden nun im Diagramm 69 die gewählten Kategorien des Items Film, anstatt für der Bewertung der Gesamtzufriedenheit, eingetragen. Da diese Variable es noch am besten vermag, mit der Gesamtzufriedenheit verglichen zu werden, sollte auch die Nummerierung weitgehend

übereinstimmen und die Flächen im Plot abgegrenzt werden. Im vorliegenden Fall, in welchem lediglich 63,4 % übereinstimmen, ist dies nur schwer erkennbar. Lediglich die ‚sehr zufrieden (1)‘-Antworten können als Gruppe visuell eingeordnet werden.

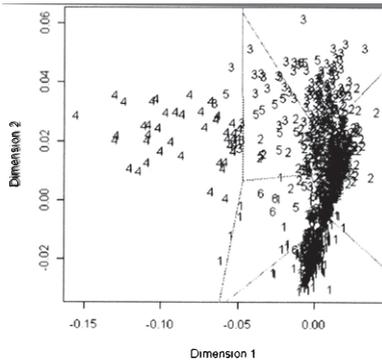


Diagramm 68: Voronoi Plot - Gesamtzufriedenheit

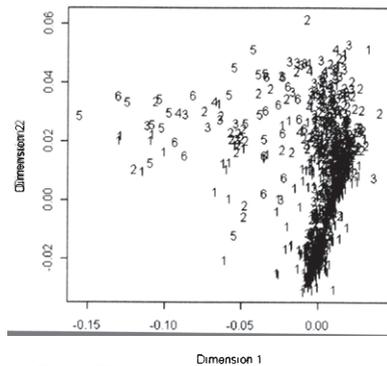


Diagramm 69: Voronoi Plot - Film

In Diagramm 70 kommt die nicht-monotone Kategorienbestimmung zum Vorschein und wird hier anhand des Items Technik/Bild/Ton beispielhaft gezeigt.

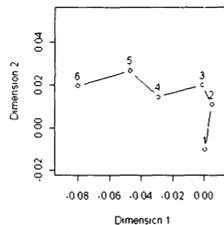


Diagramm 70: Kategorienplot - nicht-monoton – Technik/Bild/Ton

Im nächsten Beispiel wird, verglichen dazu, der Rang auf eins gesetzt, *rank-1*, und die Items als ordinal skaliert behandelt. Linke Graphik in Diagramm 71 gibt mit Sternen die Lage der Personen wieder. Die Gruppen sollten sich anhand deren Kategorien unterscheiden. Die äußersten Objektstandorte grenzen die jeweiligen Kategorien ein. Es ist jedoch erkennbar, dass eine Abgrenzung der Gesamtzufriedenheiten hier nicht gelungen ist. Die Räume der zwölf weiteren Variablen überschneiden sich noch mehr. Dies ist ein Indiz dafür, dass die zugrunde liegenden Dimensionen für die Teilzufriedenheitsvariablen noch die beste Möglichkeit sind, die Gesamtzufriedenheit darzustellen. Jede für sich ist derart voneinander unterschiedlich, dass eine Gruppeneingrenzung über die einzelnen Items keinen Sinn ergibt. Die rechte Graphik stellt anstatt der Sterne die Kategorien der Gesamtzufriedenheit dar. Es kommt deutlich zum Vorschein, dass es nicht nur Ausreißer sind, welche die Gruppengrenzen im *hullplot* derartig weitflächig ziehen, sondern allgemein ein Gruppentrend nur schwer erkennbar ist.

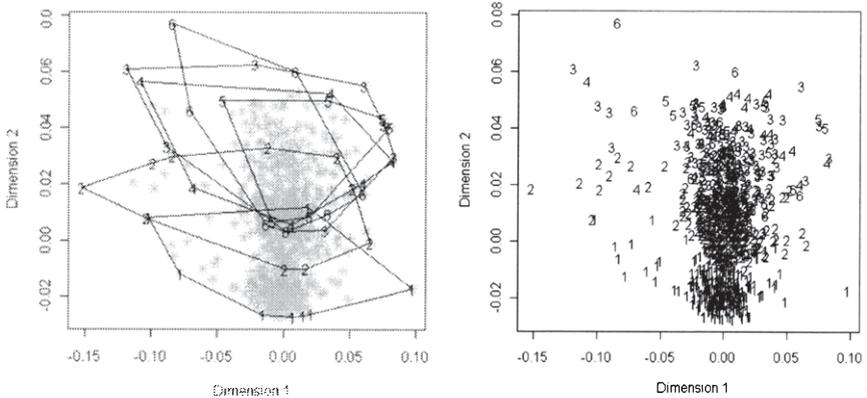


Diagramm 71: Hullplot – Labplot - Gesamtzufriedenheit

Diagramm 72 bringt für die Variable Buffet die Rangrestriktion zum Ausdruck. Die Originallage ohne Restriktion ist geknickt, die mit Rang eins gerade. Da nun Dimension 1 eine Linearkombination aus Dimension 2 darstellt, bildet die Verbindung der Buffetkategorien für das Modell mit Restriktion auch eine Gerade. Diese ist zusätzlich aufsteigend angeordnet, da in dem Modell die Restriktion ordinaler Variablen eingeführt wurde. Die unterschiedlichen Abstände zwischen den Kategorien widerspiegeln die Ordinalität der Skala.

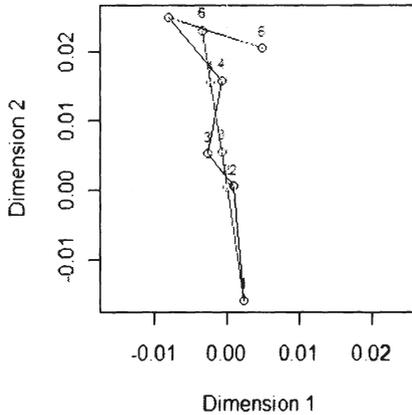


Diagramm 72: Lossplot - Buffet

Diagramm 73 vergleicht das Modell ohne Beschränkungen in der linken Graphik und das mit Beschränkungen in der rechten Graphik. Die Personenlagen sind wiederum nummeriert und die Itemkategorien in Graustufen eingezeichnet.

keine Beschränkung

Rank-1-Beschränkung

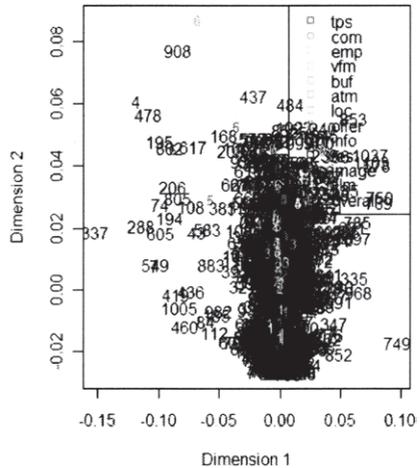
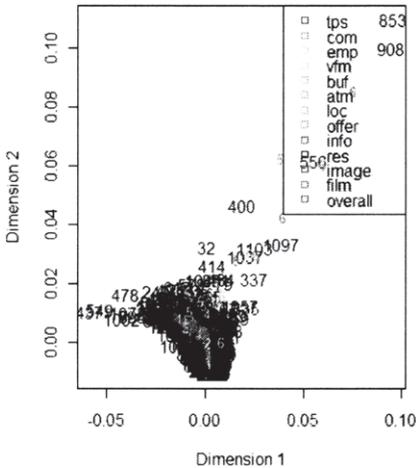


Diagramm 73: Joint Plot Vergleich

Diagramm 74 stellt obigen Unterschied für das Preis/Leistungs-Verhältnis dar. Rechte Graphik zeigt den monotonen Verlauf mit den unterschiedlichen Abständen zwischen den ordinalen Kategorien.

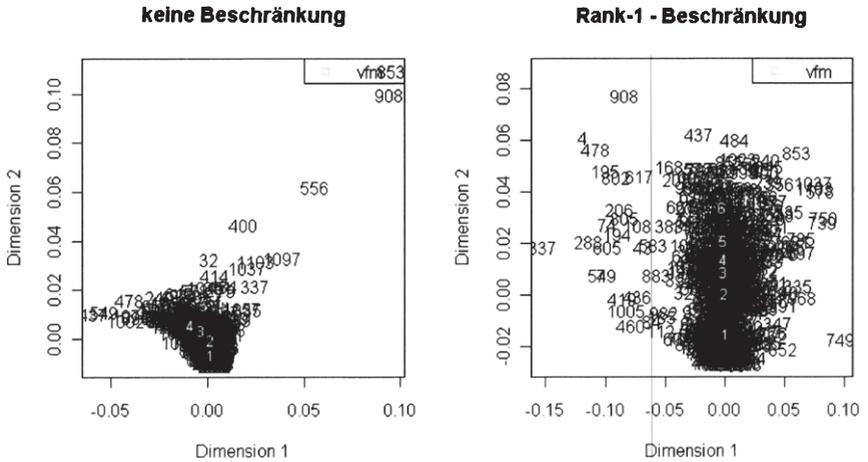


Diagramm 74: Joint Plot Vergleich - Preis/Leistung

Diagramm 75 veranschaulicht anhand der Gesamtzufriedenheit den Aspekt, warum es zuvor nicht gelungen ist, die Gruppen in den Plots auszumachen. Es können zwar die Kategorien eins und zwei durch die Dimensionen getrennt werden, die zwölf Teilzufriedenheitsitems vermögen es jedoch nicht, die niedrigeren Zufriedenheitsregionen der Gesamtzufriedenheit zu unterscheiden.

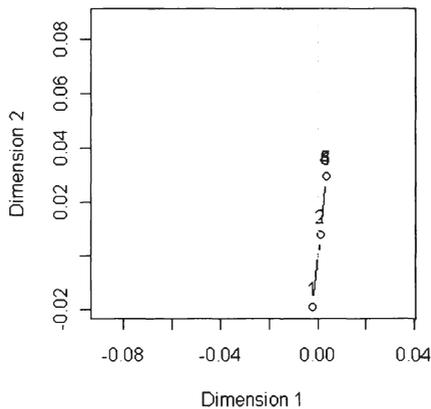


Diagramm 75: Kategorienplot - Gesamtzufriedenheit

Diagramm 76 vergleicht Unterschiede, je nachdem ob die Variablen im Zuge der Iterationsschritte der Regression als numerisch oder ordinal verwendet werden.

Es sind leichte Änderungen in den Ladungen bemerkbar. Das Skalenniveau spielt somit bei der Ermittlung der Dimensionen eine Rolle.

Lineare Regression

Monotone Regression

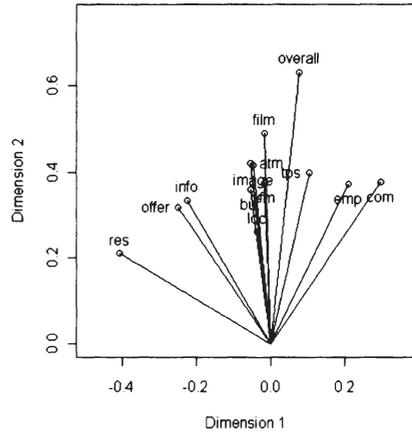
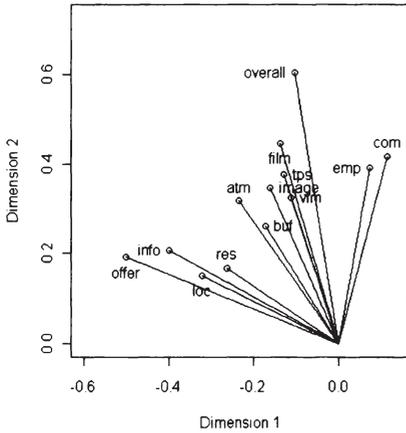


Diagramm 76: Ladungsplot - lineare vs. monotone Regression

Diagramm 77 vergleicht die originale mit der transformierten Skala sowohl für die lineare, also auch für die monotone Regression. Hier kommen die nicht-linearen Aspekte der Variablen zum Vorschein.

Lineare Regression

Monotone Regression

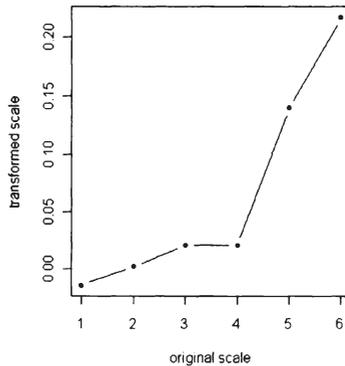
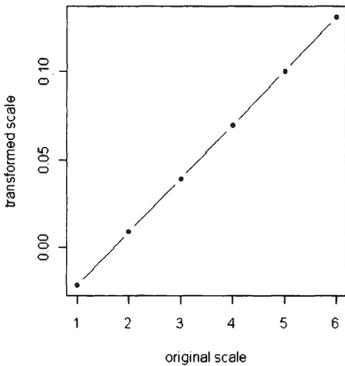


Diagramm 77: Vergleich der originalen und transformierten Skala – linear vs. monoton

Diagramm 78 visualisiert die Unterschiede der linearen Regression mit gleichen Skalenabständen und der monotonen mit unterschiedlichen Skalenabständen für das Zufriedenheitsitem Film.

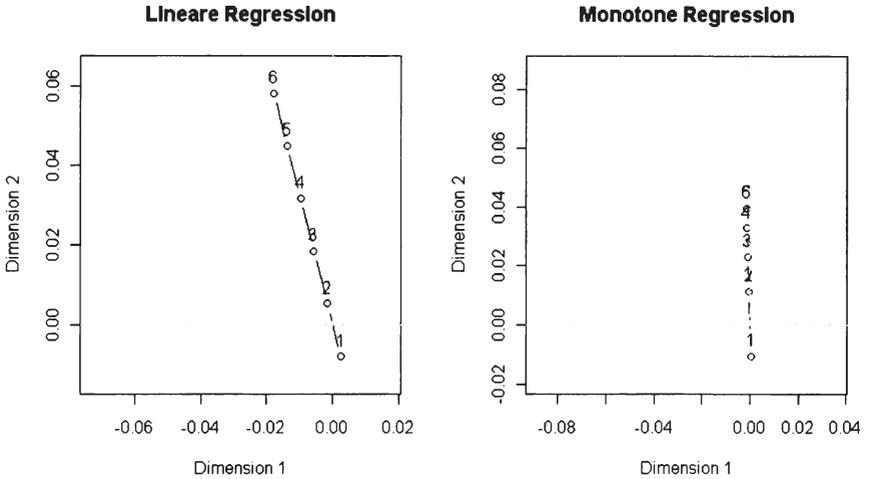


Diagramm 78: Vergleich - lineare vs. monotone Regression

Diagramm 79 zeigt eine weitere Visualisierung mit Rang-eins-Restriktion und der Behandlung der zwölf Teilzufriedenheitsitems als ordinal skalierte Variable. Darin sind die original gewählten Kategorien für Komfort auf den beiden Dimensionen eingezeichnet. Es ist schwer, Gruppen zu finden. Allgemein kann festgehalten werden, dass sich die zufriedeneren Regionen an der unteren Seite der Graphiken versammeln.

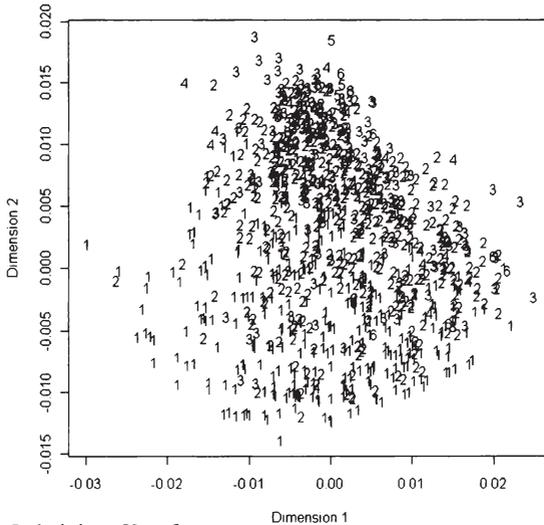


Diagramm 79: Labelplot - Komfort

Diagramm 80 zeigt die Unterschiede zwischen dem Modell mit Rang-eins-Beschränkung und dem ohne Beschränkung anhand von 20 Datenzeilen. In der linken Graphik geben die Nummerierungen auf der horizontalen durchgehenden Geraden die Kategorienwerte für das Modell mit Restriktion an. Die Kreise mit den zugehörigen Nummerierungen am Ende der Linien, welche im 45°-Winkel zur durchgehenden Geraden stehen, stehen für die Originalwerte ohne Restriktion. Je weiter entfernt die Kategorien mit deren Projektionen auf der Geraden auftreffen, desto unterschiedlicher sind die Ergebnisse der beiden Modelle. In der rechten Graphik gilt die gleiche Interpretation für die Länge der Abstände (Linien im 45°-Winkel), von deren kategorial zugehörigen orthogonalen Geraden (vertikale durchgehende) Linien. Je größer die Abstände, desto eher ist das restringierte sparsame Modell zu verwerfen.

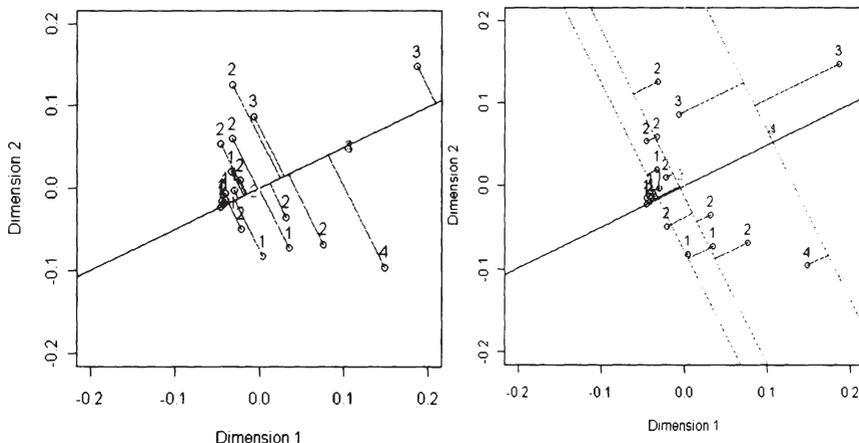


Diagramm 80: Vektorplot und Projektionsplot

Zusammengefasst wurde in diesem Kapitel gezeigt, dass die 6-Punkt-Likert Skala aufgrund ihrer Breite problembehaftet ist. Wie auch im IRT-Kapitel können die Bewertungen drei, vier, und fünf nicht sinnvoll beurteilt werden. Die Erkenntnisse dieser beiden Kapitel dienen hier als Hinweis, die Ergebnisinterpretationen der folgenden Methoden mit Vorsicht zu genießen. Weder im IRT-Kapitel konnte die Rasch-Tauglichkeit der Daten gezeigt werden, noch im Kapitel der multiplen Korrespondenzanalyse zumindest eine monotone Steigung. Somit können diese beiden Verfahren nicht zur Reskalierung der Daten verwendet werden.

4.3.4. *Multilevel Modelling: vom fixed effects model zum multilevel growth curve model*

Ziele:

Es wird untersucht, ob es standortbezogen unterschiedliche Einflüsse der Teilzufriedenheitsitems auf die Gesamtzufriedenheit mit dem Kinobesuch gibt. In den longitudinalen Modellen werden mehrere Ebenen zur Verbesserung der Bestimmung der Gesamtzufriedenheit über die Zeit hinweg eingeführt.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den 12 Monaten Oktober 2006 bis September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Bitte machen Sie folgende Angaben zu Ihrem Kinobesuch: Gesamtzufriedenheit (gemessen in den 12 Monaten Oktober 2006 bis September 2007);

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden);

Bitte machen Sie folgende Angaben zu Ihrem Kinobesuch: Kino (gemessen in den 12 Monaten Oktober 2006 bis September 2007);

Skalierung: 53 unterschiedliche Kinostandorte (nominal);

Programm:

MLwiN 2.02;

Modell:

Luke (2004) bietet einen Überblick zu den *multilevel models* (MLM), auch *hierarchical linear models*, *random coefficient models* oder *mixed effects models* genannt. Es existieren zwei Lösungsansätze, nämlich die multiple Regression und *structural equation models* (SEM). Das *ordinary least squares* (OLS) Regressionsmodell verletzt Annahmen der klassischen Regression sobald Multileveleffekte auftreten. Wird das OLS-Verfahren für Daten mit korrelierten Fehlerstrukturen verwendet, sind die Fehlerwerte kleiner als sie sein sollten. In MLM werden die brauchbaren Teile der Fehler berücksichtigt. In den Programmen HLM und R/S-Plus (package: nlme, library: mixed-effects) wird ein ML-Schätzer verwendet, um Probleme wie korrelierte Fehlerwerte zu

umgehen, indem die Unabhängigkeitsannahme aufgeweicht und korrelierte Fehlerstrukturen erlaubt werden. Probleme, wie die in longitudinalen Daten auftretende Autokorrelation, können ebenfalls berücksichtigt werden. In MLM beeinflussen Effekte höherer Levels, die auf niedrigeren Levels. Es werden Einflüsse auf aggregierter Ebene, strukturelle Gegebenheiten auf der Beziehungsebene und globale Eigenheiten, die nicht aus dem Individuum herrühren, berücksichtigt. Die Betrachtung auf disaggregierter Ebene weist die nicht im Modell enthaltene Information als Fehler des Individuums aus. Dies führt zu korrelierten Fehlern, was die Annahmen der multiplen Regression verletzt. Wird der gruppenspezifische Kontext nämlich ausgegrenzt, werden Individuen in unterschiedlichen Umgebungen gleich behandelt. Dies kann nicht gerechtfertigt werden. Ein *fixed effect* -Einfluss kann unter dem Umstand der zufälligen Variabilität nicht gewährleistet werden. MLM bieten zusätzlich besseren Umgang mit fehlenden Werten.

In den MLM gibt es deshalb einen *fixed effect* und einen *random effect*. Diese werden meist in mehreren Gleichungen angeschrieben, also in einem System an Gleichungen oder alle in die Level-1-Ebene implementiert, *mixed-effects*-Modell genannt. Tabelle 44 listet die drei Basismodelle auf.

Das *unconstrained*-Modell, auch Null-Modell genannt, hat keine Level-1- oder Level-2-Prädiktoren und wird meist als Startpunkt für komplexere Modelle verwendet. Es ist trotzdem ein MLM und liefert das gleiche Ergebnis wie das *one-way random effects*-ANOVA-Modell, also einen Gesamtmittelwert über alle Probanden, eine Variabilität zwischen den Level-2-Gruppen und eine Variabilität innerhalb der Gruppen. Das *random intercept*-Modell geht von Level-1-*intercepts* aus, welche über Level 2 variieren. Im *random intercepts- and slopes*-Modell wird diese Variation zusätzlich für die *slopes* aufgenommen. Auch ein *fixed intercept random slope*-Modell ist denkbar, beispielsweise für den Vergleich einer Kontrollgruppe und einer *treatment*-Gruppe, welche mit gleichen Startwerten nach unterschiedlicher Behandlung einen unterschiedlichen Einfluss widerfahren.

Zur Schätzung wird ML oder *restricted*-ML (REML) verwendet. Die *fixed effects* werden in beiden Methoden gleich geschätzt, die *random effects* sind bei REML weniger verzerrt. Wenn die Anzahl an Level-2-Einheiten 30 oder mehr übersteigt werden die Unterschiede in den Ergebnissen verschwindend klein. ML hat den Vorteil der möglichen Verwendung der *deviance*-Statistik für den Vergleich zwischen *fixed* und *random effects* zweier Modelle. Deshalb wird REML nur bei einer geringen Anzahl an Level 2 Gruppen empfohlen.

Class	System of Equations Model	Mixed-Effects Model	Description	Notes
1. Unconstrained	L1: $Y_{ij} = \beta_{0j} + r_{ij}$ L2: $\beta_{0j} = \gamma_{00} + u_{0j}$	$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$	One-way random effects ANOVA	Often used as a null model to estimate between-groups effects with an ICC
2. Random Intercepts	L1: $Y_{ij} = \beta_{0j} + r_{ij}$ L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$	$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + u_{0j} + r_{ij}$	Means as outcomes	Here the emphasis is on L2 predictors
	L1: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ L2: $\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10}$	$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{0j} + r_{ij}$	One-way random effects ANOVA	
3. Random Intercepts and slopes	L1: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ L2: $\beta_{0j} = \gamma_{00} + u_{0j}$ $\beta_{1j} = \gamma_{10} + u_{1j}$	$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + u_{0j} + u_{1j}X_{ij} + r_{ij}$	Random coefficients regression model	Intercepts and slopes of L1 model are allowed to vary across L2 units, but we are not modeling that variability with L2 predictors
	L1: $Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$ L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$	$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{ij} + \gamma_{11}W_jX_{ij} + u_{0j} + u_{1j}X_{ij} + r_{ij}$	Intercepts and slopes as outcomes	Level-1 intercept and slopes are modeled using Level-2 predictor(s). Note the cross-level interaction component: $(\gamma_{11}W_jX_{ij})$

Tabelle 44: Multilevel Modelle

Zur Bestimmung, ob die *fixed effects*-Parameter signifikant sind wird die *t-ratio*, ein χ^2 -Quadrat Test der Residuen unter der Annahme, dass diese nicht normalverteilt sein müssen, oder der *ML-Wald-Test* verwendet, welcher einer *Z-statistic* mit Standardnormalverteilung folgt. Zur Bestimmung des Model-Fit kann die *ML-statistic* nicht herangezogen werden, jedoch zum Modellvergleich, wenn die Modelle mit den gleichen Daten gerechnet wurden und eines der Modelle ein Teilmodell mit weniger Parametern des anderen ist. Sie gibt an, wie nahe das Modell an die Daten herankommt. Weitere Modellvergleiche werden über *goodness-of-fit*-Messungen, Devianz, AIC oder BIC durchgeführt. Durch Multiplikation des natürlichen Logarithmus der *likelihood* mit -2 kann die *likelihood* als *deviance* transformiert werden. Die Differenz der *deviance* ist χ^2 -Quadrat verteilt, wobei der Unterschied in den Freiheitsgraden dem Unterschied an geschätzten Parametern entspricht. Je geringer die *deviance* desto besser der Fit. Ist die Differenz nicht signifikant, wird das komplexere Modell verworfen. Für genestete Modelle hat das Modell mit der höheren Anzahl an Parametern immer den niedrigeren Wert. Dies stellt einen Nachteil dieser Berechnung. Da das Ziel auch Sparsamkeit ist, beziehen AIC und BIC Bestrafungen für die Anzahl an geschätzten Parametern in Formel 16 mit ein.

$$AIC = -2LL + 2p$$

$$BIC = -2LL + p \ln(N)$$

Formel 16: AIC und BIC

p ist die Anzahl an Parametern, n die Samplegröße, wobei bei letzterer jene aus Level 1 empfohlen wird. AIC und BIC ermöglichen zusätzlich die Vergleichbarkeit nicht genesteter Modelle. AIC ist für kleine Samplegrößen nicht reliabel, da meist ein *penalized quasi-likelihood* (PQL) Schätzverfahren verwendet wird, welches eine asymptotische Annäherung an die *likelihood* darstellt.

Anleitungen zu den folgenden Modelltests finden sich in Luke (2004). Für den Modell-Fit der OLS wird normalerweise das R^2 berechnet, welches den %-anteil der durch die Prädiktoren erklärten Varianz angibt. Im Mehrebenenmodell gibt es für jeden Level ein eigenes R^2 . Die Einführung zusätzlicher Parameter kann zu einem kleineren oder sogar negativen R^2 führen. Im Mehrebenenmodell wird das R^2 als proportionale Reduktion des vorhergesagten Fehlers interpretiert. Für den Fehler auf Level 1 bedeutet dies, Fitverbesserung der individuellen Werte, für Level 2 auf Gruppenebene. Zur Verfügung stehen graphische Modelltests wie der *boxplot* der Residuen für die Gruppen, ein *scatterplot* der standardisierten Residuen gegen die gefitteten Werte um Heteroskedastizität zu prüfen, der *normal quantile-quantile plot* bzw. *QQ-plot* der Level-1-Residuen, um die

Normalität der Daten zu prüfen, ein Plot der Level 2 *random effects*, um die Normalverteilung mit Mittelwert null zu prüfen, oder eine *scatterplot*-Matrix, um die *random effects* durch Betrachtung der Korrelation auf Unabhängigkeit und Verteilung um null zu untersuchen. Weitere Modellbewertungen stellen die Prüfungen der Unabhängigkeit und Normalverteilung der Level-1-Residuen, oder die Normalverteilung und Unabhängigkeit der Zufallseffekte zwischen den Gruppen dar. Auch der *random*-Teil, also die Variabilität zwischen Level 2 Einheiten auf *intercepts* und *slopes* des Level 1 Modells, kann mit Hilfe der *empirical Bayes estimation* untersucht werden. Je weniger reliabel die prognostizieren Werte der Abhängigen für die Gruppen sind, desto eher empfiehlt sich die Verwendung des Gesamtmittelwertes als Schätzung für die Gruppen.

Weitere zu beachtende Punkte betreffen *centering* bzw. *reparameterization* der Level-1-Prädiktoren, wenn es einen *random slope* gibt. Diese können Teile des Modells und die Interpretation aufgrund der Transformierung der Skala ändern. Es besteht die Möglichkeit einer Zentrierung über den Gesamtmittelwert, *grand mean*, oder auch über den Gruppenmittelwert. Letzterer wird häufig für Wachstumskurvenmodelle, *growth curve modeling*, longitudinaler Daten verwendet.

Luke (2004) erwähnt Erweiterungen, wie beispielsweise *generalized hierarchical linear models* (GHLM) für nicht-kontinuierliche oder nicht-normalverteilte Daten, also binäre Daten, *proportions*, Zähldaten oder ordinale abhängige Variablen, welche die Normalitätsannahmen und Homoskedastizitätsannahmen der Fehler verletzen. Auch Level-3-Modelle, bei welchen die Level-1-*intercepts* und -*slopes* zufällig auf Level 2 und 3 variieren können, sind denkbar.

Auch ein Kapitel zu den später folgenden longitudinalen Untersuchungen ist angeführt. Dabei werden die Zeitpunkte als genestet über die Personen in der hierarchischen Betrachtungsweise gesehen. Auch zusätzliche Kovariate können verarbeitet werden. Die *repeated-measures*-MANOVA weist Probleme bei der Analyse von ungleichmäßigen Beobachtungen bzw. regelmäßigen Beobachtungen mit Fehlerwerten auf. Somit ist diese nicht zu empfehlen. MLM können die Daten ohne Personen- oder Informationsverlust verarbeiten und erlauben *time slopes* über Personen zu variieren. Diese Variation kann zusätzlich über Prädiktoren, welche den Personenlevel beeinflussen, erklärt werden. Für einen Modellvergleich wird BIC empfohlen, vor allem bei einer großen Samplegröße, da diese zur Berechnung miteinbezogen wird. Nicht longitudinale

Studien gehen von Normalverteilung und Unabhängigkeit der Fehler aus. Diese Annahmen können für longitudinale Analysen oft nicht vorausgesetzt werden und es muss eine Fehlerkovarianzstruktur vordefiniert werden. Die bekannteste ist die *unrestricted-* bzw. *unstructured-*Fehlerstruktur, welche keine Annahmen beinhaltet und somit jede erdenkliche Fehlerkorrelation über die Beobachtungen zulässt. Dieses Modell weist die höchste Anzahl an *random-*Parametern auf, da jeder *lag* eine eigene geschätzte Kovarianz besitzt. Es ist das am wenigsten sparsame Modell und wird oft als Ausgangsmodell zu weiteren Modellvergleichen herangezogen. Aber auch konstante Kovarianzen über alle Wellen können im *homogeneous error-*Modell angenommen werden. Es ist das sparsamste und beschreibt die selbigen, aber in der Realität kaum rechtfertigbaren Annahmen, welche auch hinter der *repeated-measures univariate-ANOVA* unter dem Begriff *compound symmetry* bekannt sind. Auch eine autoregressive Struktur z.B. erster Ordnung, *first-order autoregressive structure*, ist denkbar. Die Fehlerterme sind dann über *first-order lags* korreliert. Die Korrelation zwischen zwei benachbarten Zeitpunkten ist folglich immer gleich. Da eine zeitliche Veränderung möglich ist, erlaubt das Modell über geringere Korrelation weiter entfernter Zeitpunkte erweiterten Spielraum. Im Vergleich zum *homogeneous error-*Modell wird ein zusätzlicher Parameter, die Korrelation erster Ordnung ρ , geschätzt.

4.3.4.1. *Random intercept fixed slope model* - Kinodaten

Die Abhandlung dieses Kapitels folgt grob dem Aufbau des User's Guide for MLwiN von Rasbash et al. (2005). Es existieren zwei Levels, die Probanden auf Level 1 und Kinostandorte auf Level 2, wobei der Zusammenhang in die Berechnung miteinbezogen wird. In den einzelnen Gruppen können unterschiedliche Effekte entstehen. Im Zuge herkömmlicher Clusterverfahren sind die Personen innerhalb eines Clusters voneinander abhängig. Die Cluster selbst sind voneinander unabhängig. Diese Unabhängigkeit wird normalerweise nicht berücksichtigt. Die Zufallsvarianz wird folglich unterschätzt, da zwei unterschiedliche Gruppen, verglichen mit nur einer Stichprobe, eine größere zufällige Streuung besitzen können. MLM werten einzelne Gruppen deshalb als unabhängig. Es werden einerseits Abweichungen zum Gesamtmittelwert aller Kinostandorte in die Berechnung miteinbezogen, andererseits auch Abweichungen innerhalb eines Standortes zu dessen Mittelwert. So wird die Varianz zwischen und innerhalb der einzelnen Kinos berücksichtigt. Beide Varianzarten zusammen bilden die Gesamtvarianz. Je höher der Anteil der Varianz auf Kinostandortebene an der Gesamtvarianz, desto unterschiedlicher sind die Standorte bzw. desto ähnlicher die Probanden innerhalb eines Standortes. Die Streuung der einzelnen Cluster und die Form der Verteilung

innerhalb der Cluster werden nicht beachtet. Unterschiede ergeben sich lediglich durch die unterschiedlichen Mittelwerte der Cluster. Ein Grund für die Verwendung von MLM liegt auch in der Schätzung wesentlich weniger Parameter, welche für die einzelnen Kinostandorte geschätzt werden müssten. Auch die Standardabweichungen der Regressionskoeffizienten würden ohne Gruppenannahme unterschätzt werden. Die Kinostandorte werden als Zufallsstichproben einer zugrunde liegenden Population gewertet. Die OLS-Technik wird erweitert und die Variation zwischen den Standorten der Stichprobe dazu genutzt, Aussagen zur Variation in der Grundgesamtheit zu machen.

In der Formelstruktur $y_{ij} = a + bx_{ij} + u_j + e_{ij}$ stellt $a + bx_{ij}$ den fixen Teil dar. i steht für die Variation auf Besucherebene und j für die Variation auf Standortebene. a_j besteht aus $a + u_j$, nämlich ein fixer *intercept* a mit einer standortabhängigen Variation u_j . u_j und e_{ij} beschreiben die zufälligen Teile mit Mittelwert null. Es gibt in diesem Beispiel nur eine erklärende Variable und eine abhängige. a und b , der *intercept* und *slope* sind fix und werden ebenfalls geschätzt. Ebenso werden die Varianzen der zufälligen Teile σ^2_u und σ^2_e geschätzt. Die einzigen zufälligen Parameter sind die *intercept*-Varianzen auf Ebene der Besucher und auf Ebene der Standorte. Dieses Modell wird auch *variance component*-Modell genannt. Um das Modell für komplexere Beschreibungen verständlicher zu machen, wird eine erklärende Variable x_0 durchgehend mit dem Wert eins für alle Besucher eingefügt. β_0, β_1 , usw. stehen für die fixen Parameter und die Indizes 1, 2, usw. geben die Nummerierung der erklärenden Variablen an. Die Zufallsparameter erhalten den Index 0. Daraus ergibt sich die Gleichung $y_{ij} = \beta_0 x_0 + \beta_1 x_{1ij} + u_{0j} x_0 + e_{0ij} x_0$. Die Koeffizienten ergeben sich aus $y_{ij} = \beta_{0ij} x_0 + \beta_1 x_{1ij}$ mit zugehörigen Abweichungen $\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$. Die zufällige Variation der abhängigen Variablen, wird durch die zufälligen Koeffizienten der erklärenden Variablen beschrieben. Der Koeffizient x_0 ist zufällig auf Level 1 und 2. Der Index 0 der Level-1- und -2-Zufallsvariablen bedeutet, dass diese zu x_0 gehören. Die abhängige Variable wird vorerst als normalverteilt und intervallskaliert behandelt $y \sim N(XB, \Omega)$. Ω steht für die Varianzen und Kovarianzen der Zufallsterme über die einzelnen Levels. XB beschreibt den fixen Part des Modells,

$$\begin{pmatrix} \beta_0 x_{011} + \beta_1 x_{111} \\ \beta_0 x_{021} + \beta_1 x_{121} \\ \dots\dots\dots + \dots\dots\dots \\ \dots\dots\dots + \dots\dots\dots \end{pmatrix}.$$

Im Kinobeispiel werden zur Erklärung der Varianz zwei Prädiktoren aufgenommen, die Bewertungen der Zufriedenheit auf Level 1 und die

Kinostandorte auf Level 2. Der Bewertungszeitpunkt spielt vorerst keine Rolle. Es werden alle Kinobesuchsbewertungen der zwölf Monate verwendet. Die Fallzahl wird in den jeweiligen Modellberechnungen angeführt. Zuerst wird in Modell 1 untersucht, innerhalb welcher Variablen es signifikante Unterschiede der *intercepts* auf Level 2 Ebene in Bezug auf die Gesamtzufriedenheit gibt, indem eine Regression unter Miteinbeziehung des Kinostandortlevels durchgeführt wird. Zur Interpretation der Prädiktoren wird die Division der Koeffizienten durch deren jeweils in Klammer zugehörigen Standardabweichungen herangezogen. Diese müssen bei einem Signifikanzniveau von 95 % über dem Wert 1,96 der T-Verteilung liegen.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij}, cons$$

$$\beta_{0ij} = 5,013(0,027) + u_{0i} + e_{0ij}$$

$$[u_{0i}] \sim N(0, \Omega_u) : \Omega_u = [0,012(0,005)]$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [1,511(0,017)]$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 54471,160 (16747 \text{ of } 16747 \text{ cases in use})$$

Modell 1: Interceptvariation - Gesamtzufriedenheit

Tabelle 45 listet die Ergebnisse für die Gesamt- und Teilzufriedenheitsitems auf. Hier muss angemerkt werden, dass die Level-1-Varianzen, verglichen mit den Level 2 Varianzen, in allen Zufriedenheitsvariablen deutlich überwiegen.

Item	Koeffizient	Standardabweichung
Gesamtzufriedenheit	0,012	0,005
Technik/Bild/Ton	0,090	0,024
Komfort	0,128	0,033
Servicepersonal	0,026	0,009
Preis/Leistung	0,070	0,021
Buffet	0,045	0,015
Atmosphäre	0,068	0,020
Standort/Erreichbarkeit	0,038	0,012
Filmangebot	0,088	0,024
Informationsangebot	0,065	0,018
Reservierungsmöglichkeiten	0,062	0,018
Erscheinungsbild/Image	0,075	0,021

Tabelle 45: Interceptvariation – Gesamt- und Teilzufriedenheiten

Signifikante Unterschiede auf Kinostandortebene, Level 2, zeigen sich in allen Items, sofern deren Einflüsse einzeln betrachtet werden. Nun wird untersucht,

wie sich die Teilzufriedenheitsitems auf Kinostandortebene in Bezug zur Gesamtzufriedenheit verhalten. Modell 2, ohne geschätzte Werte, dient zuvor nochmals der detaillierten Symbolerklärung.

$$\begin{aligned}
 y_{ij} &\sim N(XB, \Omega) \\
 y_{ij} &= \beta_{0ij}x_0 + \beta_1x_{ij} \\
 \beta_{0ij} &= \beta_0 + u_{0j} + e_{0ij} \\
 [u_{0j}] &\sim N(0, \Omega_u): \Omega_u = [\sigma_{u_0}^2] \\
 [e_{0ij}] &\sim N(0, \Omega_e): \Omega_e = [\sigma_{e_0}^2]
 \end{aligned}$$

Modell 2: Teilzufriedenheiten und Gesamtzufriedenheit

- XB = Mittelwert, durch den fixen Teil spezifiziert
- Ω = zufälliger Teil, durch Kovarianzmatrix Ω beschrieben, beinhaltet die Kovarianzmatrizen aller Levels
- y = Gesamtzufriedenheit
- x = Zufriedenheit mit Teilzufriedenheitsitem
- i = Index auf Level 1 (Personenebene)
- j = Index auf Level 2 (Standortebene)
- x_0 = fixer Parameter (durchgehend 1) für den *intercept*, zufällig auf Besucher und Standortebene
- β_{0ij} = Regressionsparameter für den *intercept*, für die Standorte unterschiedlich
- β_1 = Regressionsparameter für den *slope*, für alle Standorte einheitlich
- u_{0j} = Level 2 Residuen, Wert um welchen der Standort um den allgemeinen *intercept* variiert
- e_{0ij} = Level 1 Residuen, Wert um welchen die Besucher um den Standortmittelwert variieren
- $\sigma_{u_0}^2$ = Varianz des Standortlevel-Zufalleffektes, normalverteilte Zufallsabweichungen auf Standortebene mit Mittelwert null
- $\sigma_{e_0}^2$ = Varianz des Besucherlevel-Zufalleffektes, normalverteilte Zufallsabweichungen auf Besucherebene mit Mittelwert null

In Modell 3 werden die Ergebnisse für Technik/Bild/Ton eingefügt. Geschätzt wird mit iterative *generalized least squares* (IGLS). Das Modell konvergiert, wenn die Änderung sämtlicher Parameter kleiner 10^{-2} ist.

Ändert sich die Technik/Bild/Ton-Zufriedenheitsbewertung, wird bei Vorliegen von Signifikanz der Wert des zugehörigen Prädiktors zur Gesamtzufriedenheit hinzugezählt bzw. abgezogen. Die Steigung ist für alle Kinostandorte gleich, 0,388, und mit einer Standardabweichung von 0,011 hoch signifikant. Die

intercepts betreffend gibt es Unterschiede. Ihr Mittelwert beträgt 2,977 mit einer Standardabweichung von 0,064. Die *intercepts* der jeweiligen Standorte, welche den Level 2 Residuen entsprechen, schwanken signifikant um ihren Mittelwert mit 0,016 und einer zugehörigen Standardabweichung von 0,006. Die Personenbewertungen schwanken um ihren Standortmittelwert mit 1,404 und zugehöriger Standardabweichung von 0,015. Diese sind deutlich höher, verglichen mit den Schwankungen auf Level 2. Die durchschnittliche Abweichung einer Person beträgt $0,016+1,404=1,420$. Die standortbezogene Abweichung stellt somit einen Anteil von $0,016/1,404=1,13\%$ dar, die Intra-Standort-Korrelation. Dies stellt ein Ähnlichkeitsmaß für Besucher innerhalb eines Standortes im Vergleich zu anderen Standorten dar. Je höher sie ist, desto größer ist der Unterschied zwischen den Standorten.

$$\begin{aligned}
 ges_{resp.kino} &\sim N(XB, \Omega) \\
 ges_{resp.kino} &= \beta_{0resp.kino}cons + 0,388(0,011)tbt_{resp.kino} \\
 \beta_{0resp.kino} &= 2,977(0,064) + u_{0kino} + e_{0resp.kino} \\
 [u_{0kino}] &\sim N(0, \Omega_u) : \Omega_u = [0,016(0,006)] \\
 [e_{0resp.kino}] &\sim N(0, \Omega_e) : \Omega_e = [1,404(0,015)] \\
 -2 * \text{loglikelihood (IGLS Deviance)} &= 53245,550(16747 \text{ of } 16747 \text{ cases in use})
 \end{aligned}$$

Modell 3: Technik/Bild/Ton-Zufriedenheit bezogen auf Gesamtzufriedenheit

Diagramm 81 veranschaulicht das *random intercept*-Modell. Die dicke Linie in der Mitte aller Linien zeigt die vorhergesagte Schätzung der fixen Parameter ohne die standortspezifische Variation des *intercepts* ($\hat{y}_{ij} = \hat{a}_0cons + \hat{a}_1tbt_{ij}$). Die dünneren Linien zeigen die vorhergesagten Regressionsgeraden aller Standorte unter Miteinbeziehung des Parameters u_{0j} ($\hat{y}_{ij} = \hat{a}_{0j}cons + \hat{a}_1tbt_{ij}$). Der *slope*-Koeffizient β_1 ist für alle Standorte gleich. Die Standortlevelresiduen modifizieren lediglich den *intercept* β_0 . Der Unterschied zwischen den Standorten bleibt über die gesamte Spanne der Gesamtzufriedenheit konstant und nennt sich aufgrund dieser gleichbleibenden Veränderung *simple level 2 variation*. Es gibt Standorte, welche bei gleicher Technik/Bild/Ton-Zufriedenheit unterschiedliche Gesamtzufriedenheiten aufweisen. Die Abszissenachse zeigt die Technik/Bild/Ton-Zufriedenheit auf einer Skala von ‚sehr zufrieden (6)‘ bis ‚nicht zufrieden (1)‘. Diese stellen die Spanne der ursprünglichen Antworten dar. Die Ordinatenachse gibt die prognostizierten Werte an.

Hier ändert sich die Stärke der Auswirkung der Technik/Bild/Ton-Zufriedenheitsunterschiede auf die Gesamtzufriedenheit über deren Spanne hinweg nicht. Deshalb ergeben sich standortspezifisch parallele Verläufe.

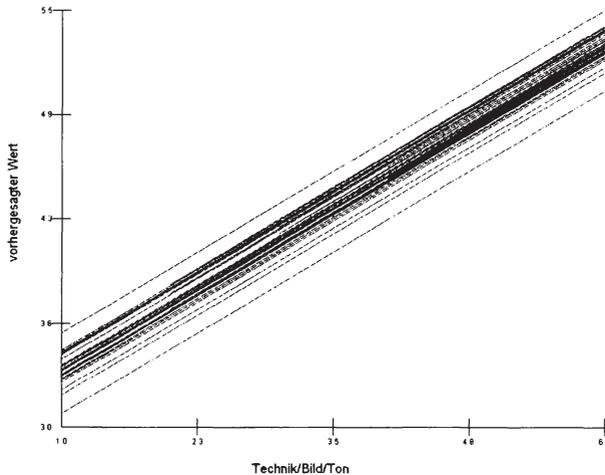


Diagramm 81: Random Intercept-Modell - Technik/Bild/Ton

4.3.4.2. *Random intercept random slope model* - Kinodaten

Es stellt sich aber die Frage, ob die Technik/Bild/Ton-Zufriedenheitssprünge zu standortspezifischen unterschiedlichen Veränderungen der Gesamtzufriedenheit führen. Im Modell 4 wird ein *random slope*-Modell geschätzt. Zusätzliche standortspezifische Abweichungen den *slope* betreffend werden für β_1 mitgeschätzt. Der zuvor fixierte Parameter für die *slope*-Werte ist somit von Standort zu Standort, also auf Level 2 Ebene unterschiedlich.

β_0 und β_1 sind die fixen Regressionskoeffizienten, welche die durchschnittliche Gerade über alle Standorte bestimmen. Die zugehörigen standortspezifischen Zufallsabweichungen auf Level 2 sind mit u bezeichnet, die besucher-spezifischen Zufallsabweichungen auf Level 1 mit e . u_{0j} und u_{1j} sind die standortspezifischen Abweichungen des *intercepts* und *slopes*. Sie sind normalverteilt mit Mittelwert null und Kovarianzmatrix Ω_u . Es gibt zwei Zufallsvariablen auf Level 2. Folglich besteht die Kovarianzmatrix aus σ^2_{u0} , die Abweichungen auf Standortebene den *intercept* betreffend, σ^2_{u1} , die Abweichungen auf Standortebene den *slope* betreffend, und σ^2_{u01} , die Kovarianz zwischen *slope* und *intercept*. e_{0ij} gibt die Besucherabweichungen auf Level 1 an. Der Wert von 0,392 (0,018) entspricht beinahe dem Wert im Modell mit nur einem *slope* 0,388 (0,011). Im aktuellen Modell variieren die *slope*-Werte jedoch mit einem Wert von 0,003 (0,002). Der aktuelle *intercept*-Wert beträgt 2,954 (0,093) und variiert mit 0,086 (0,055) auf Standortebene.

$$y_{ij} \sim N(XB, \Omega)$$

$$y_{ij} = \beta_{0ij}x_0 + \beta_{1j}x_{1ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [\sigma_{e0}^2]$$

$$ges_{resp.kino} \sim N(XB, \Omega)$$

$$ges_{resp.kino} = \beta_{0resp.kino}cons + \beta_{1kino}tbt_{resp.kino}$$

$$\beta_{0resp.kino} = 2,954(0,093) + u_{0kino} + e_{0resp.kino}$$

$$\beta_{1kino} = 0,392(0,018) + u_{1kino}$$

$$\begin{bmatrix} u_{0kino} \\ u_{1kino} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,086(0,055) & \\ -0,014(0,010) & 0,003(0,002) \end{bmatrix}$$

$$[e_{0resp.kino}] \sim N(0, \Omega_e) : \Omega_e = [1,402(0,015)]$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 53240,360(16747 \text{ of } 16747 \text{ cases in use})$$

Modell 4: Random Slope-Modell – Technik/Bild/Ton

Es zeigt sich eine negative Kovarianz zwischen den geschätzten *intercepts* und *slopes* -0,014 (0,010). Hohe *intercept*-Werte gehen mit niedrigen *slopes* einher, und vice versa. Es ergibt sich eine Korrelation von $\frac{-0,014}{\sqrt{0,003 * 0,086}} = -0,8716$.

Je besser ein Standort bewertet wird, desto geringer ist die Chance auf Verbesserung, je schlechter, desto höher der Grad an möglicher Verbesserung. Die einzelnen Besucher variieren um die standortspezifischen Regressionsgeraden mit einem Wert von 1,402 (0,015), die Level 1 Residuen.

Der Wert $-2 * \text{loglikelihood (IGLS Deviance)} = 53.240,360$ dient im Vergleich zum vorherigen einfachen *random intercept*-Modell zur Prüfung der Tauglichkeit und Sinnhaftigkeit der standortspezifischen *slopes* im komplexeren Modell. Der Wert im einfachen Modell liegt bei 53.245,550 und ist um 5,190 gesunken. Das komplexere Modell beinhaltet zwei zusätzliche Parameter, die Varianz der *slope*-Residuen u_{1j} und deren Kovarianz mit den *intercept*-Residuen u_{0j} . Aus der χ^2 -Verteilung folgt ein Signifikanzwert von 0,074646 und die Veränderung ist somit knapp nicht signifikant. Das komplexere Modell wird zu Gunsten des einfachen Modells verworfen. Diagramm 82 veranschaulicht die Ergebnisse. Die Ordinatennachse gibt die prognostizierten Werte durch den gruppenspezifischen *random slope* und *random intercept* an. Dazu werden Prädiktorwerte erzeugt, welche den gruppenspezifischen *intercept* und *slope* beinhalten. Jeder der Strahlen zeigt den Zusammenhang zwischen Zufriedenheit mit Technik/Bild/Ton

eines Kinostandortes und der Bewertung der Gesamtzufriedenheit an selbigem Standort, genannt *complex level 2 variation*. Einige Teilzufriedenheiten zeigen eine breitere Streuung auf der y-Achse. Dies bedeutet, dass diese in Bezug auf den *random slope* nicht auf aggregierter Ebene behandelt werden dürfen. Diese Teilzufriedenheitsvariablen haben unterschiedliche Auswirkungen auf die Gesamtzufriedenheit zwischen den jeweiligen Kinostandorten. Für Standorte mit steileren Geraden haben Änderungen der Technik/Bild/Ton-Zufriedenheiten stärkere Auswirkungen auf Änderungen in der Gesamtzufriedenheit. Für Standorte mit flachen Geraden bewirken Technik/Bild/Ton-Zufriedenheitsveränderungen weniger. Letztere Standorte sind weniger sensibel was die Technik/Bild/Ton-Bewertung betrifft. Sie haben einerseits weniger Möglichkeit durch Qualitätsverbesserungen von Technik/Bild/Ton Verbesserungen in der Gesamtzufriedenheit zu erzielen, andererseits haben Verschlechterungen mildere Auswirkungen.

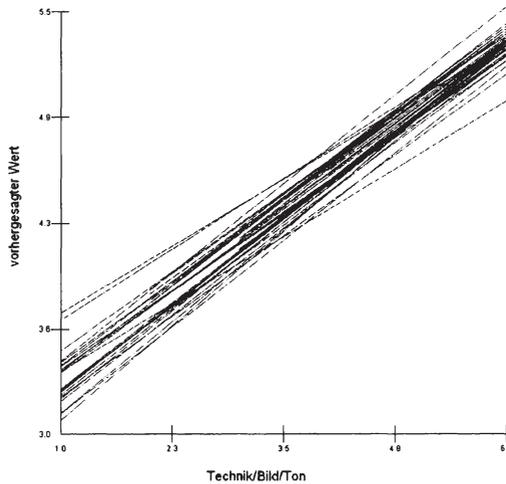


Diagramm 82: Random Slope Modell – Technik/Bild/Ton

Die χ^2 -Vergleiche in Tabelle 46 zeigen, für welche der elf Variablen es sinnvoll ist, standortspezifische *slopes* zu schätzen, signifikante Werte, und für welche nicht, nicht-signifikante Werte. Dafür werden die *random intercept*- mit den *random slope*-Modellen mittels χ^2 -Differenztest verglichen. Es zeigt sich, ob es Sinn macht, unterschiedliche Auswirkungen auf die Gesamtzufriedenheit, je nach Attributzufriedenheitslevel, mitzuberechnen, oder ob die Gesamtzufriedenheit unabhängig von der Zufriedenheit mit dem Serviceattribut auf Standortebene ist. Signifikante Werte finden sich für Komfort, Preis/Leistung, Standort, Filmangebot, Reservierungsmöglichkeiten und Erscheinungs-

bild/Image. Bei Technik/Bild/Ton, Personal, Buffet, Atmosphäre und Informationsangebot gibt es keine standortspezifischen Unterschiede, was die Auswirkung der Attributzufriedenheit auf die Gesamtzufriedenheit betrifft und ein *slope*-Wert über alle Standorte ist daher sinnvoller.

Serviceattribut	-2*log-likelihood (IGLS Deviance)		Differenzwert log-likelihood (Devianz) und Freiheitsgrade	Signifikanzwert
	ohne random slope	mit random slope		
Technik/Bild/Ton	53.245,550	53.240,360	5,190 (2)	0,0746460
Komfort	53.259,800	53.249,910	9,890 (2)	0,0071189
Personal	53.373,880	53.373,880	0 (2)	1,0000000
Preis/Leistung	53.297,020	53.290,010	7,010 (2)	0,0300470
Buffet	53.810,540	53.807,980	2,56 (2)	0,2780400
Atmosphäre	52.935,870	52.934,430	1,44 (2)	0,4867500
Standort	53.902,490	53.880,600	21,89 (2)	1,7646e ⁻⁰⁰⁵
Filmangebot	53.765,380	53.744,550	20,83 (2)	2,9979e ⁻⁰⁰⁵
Informationsangebot	53.733,480	53.728,910	4,57 (2)	0,1017700
Reservierungsmöglichkeiten	54.063,350	54.046,700	16,65 (2)	0,00024238
Erscheinungsbild/Image	53.089,130	53.081,420	7,71 (2)	0,0211740

Tabelle 46: Random Slope-Modell - Teilzufriedenheiten

4.3.4.3. *Non-linear random intercept random slope model* - Kinodaten

Um festzustellen, ob es standortbezogen unterschiedliche nicht-lineare Auswirkungen von Technik/Bild/Ton auf die Gesamtzufriedenheit gibt, wird das Quadrat der Originalantworten der Technik/Bild/Ton-Zufriedenheitsfrage als weitere erklärende Variable im Modell 5 aufgenommen. Das Modell ohne quadratischen Term weist einen *likelihood*-Wert von 53.240,360 auf. Es ergibt sich eine Devianz von 61,3 bei vier Freiheitsgraden. Die Modellverbesserung ist mit einem Signifikanzwert 0,00000 hoch signifikant, was für die Einführung eines quadratischen Terms in das Modell spricht. Diagramm 83 zeigt die prognostizierten Werte für die unterschiedlichen Technik/Bild/Ton-Zufriedenheitslevels.

Der Effekt der Technik/Bild/Ton-Zufriedenheit auf die Gesamtzufriedenheit nimmt in höheren Zufriedenheitslagen nicht-linear zu. Somit können die größten Gesamtzufriedenheitssprünge in den hohen Zufriedenheitslevels erzielt werden. Verbesserungen der Technik/Bild/Ton-Qualität haben die stärksten Effekte für

Probanden, welche bereits sehr zufrieden sind und weitere Zufriedenheitsverbesserungen erfahren. Die Zufriedenheit mit Technik/Bild/Ton wandelt sich somit über deren Skala in den höheren Bereichen zum *delight*-, *excitement*- oder Begeisterungsfaktor.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons + \beta_{1ij} tbt_{ij} + 0,019(0,008) tbt_{ij}^2$$

$$\beta_{0ij} = 3,305(0,178) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = 0,222(0,073) + u_{1j} + e_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,043(0,041) & \\ -0,007(0,007) & 0,002(0,001) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 3,132(0,352) & \\ -0,318(0,073) & 0,057(0,015) \end{bmatrix}$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 53179,060 (16747 \text{ of } 16747 \text{ cases in use})$$

Modell 5: Nicht-lineares Random Slope-Modell – Technik/Bild/Ton

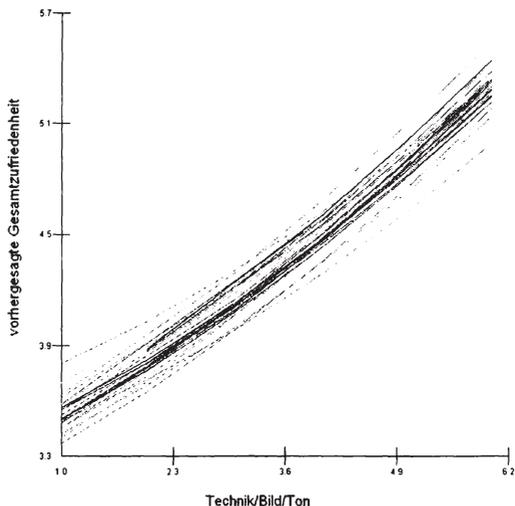


Diagramm 83: Nicht-lineares Random Slope Modell – Technik/Bild/Ton

Je größer die Level-2-Residuen im Vergleich zu den Gesamtresiduen, Level-1 + 2-Residuen, desto stärker der Standorteffekt. Um einen tieferen Einblick zu bekommen, werden in Diagramm 84 für das *random intercept*-Modell standortspezifisch die Residuen mit deren Konfidenzintervallen ausgegeben. Jene Standorte, in welchen die Konfidenzintervalle nicht die strichlierte Linie überschreiten, weichen signifikant von dem über alle Standorte prognostizierten fixen Parameterwert ab. Diese sind in der Minderheit. Höhere Werte zeigt nur

ein Standort, niedrigere einige Standorte. Weiter ist ersichtlich, dass es Standorte mit höheren und niedrigeren Konfidenzintervallen gibt. Diese entstehen hier auch aufgrund der enthaltenen Fallzahlen. Allgemein kann gesagt werden, dass eine größere Fallzahl zu kleineren Konfidenzintervallen führt, und umgekehrt.

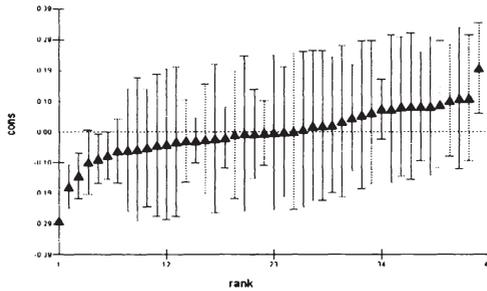


Diagramm 84: Random Intercept-Modell - Residuen und Konfidenzintervalle

Nun wird das *random slope*-Modell für Diagramm 85 herangezogen. In der ersten Graphik sind die *intercepts* gegen die *slopes* der Standorte geplottet. Dreiecke markieren den Standort mit dem niedrigsten (schwarz) und mit dem höchsten (grau) *slope*-Wert, bei etwa gleichen *intercept*-Werten. In der zweiten Graphik werden diese in die Strahlenwolke eingezeichnet, um die Lage zu bestimmen. In der dritten Graphik werden lediglich die beiden markierten Standorte mit deren 95%-igen Konfidenzintervallen und die durchschnittliche Regressionsgerade über alle Standorte, die dicke Gerade, eingezeichnet.

Die *intercepts* der beiden markierten Standorte liegen leicht über dem durchschnittlichen *intercept* aller Standorte und besitzen in etwa den gleichen Wert. Dies ist auch aus zweiter Graphik ersichtlich, in welcher die Regressionsgeraden beider Standorte in die Geraden aller Standorte eingezeichnet sind. Die graue Gerade zeigt, vergleichen mit der schwarzen, in der ersten Graphik den höheren *slope*, was in zweiter Graphik durch die stärker ansteigende Gerade angezeigt wird. Die Gesamtzufriedenheit wird, die beiden Standorte vergleichend, in den höheren Zufriedenheitslevels, mit dem Serviceleistungsaspekt Technik/Bild/Ton, stärker ausdifferenziert. Die 95%-igen Konfidenzintervalle in Graphik drei überschneiden sich in den höheren Werten nicht mehr, in den niedrigeren Werten treffen diese beinahe gänzlich aufeinander. Dafür ist der standortspezifische *slope*-Effekt verantwortlich. Die Konfidenzintervalle sind in den schlechteren Bewertungen (1) größer, verglichen mit den besseren (6). Es wird auf eine differenziertere Technik/Bild/Ton-Auswirkung auf die Gesamtzufriedenheit in niedrigeren Zufriedenheitslevels geschlossen.

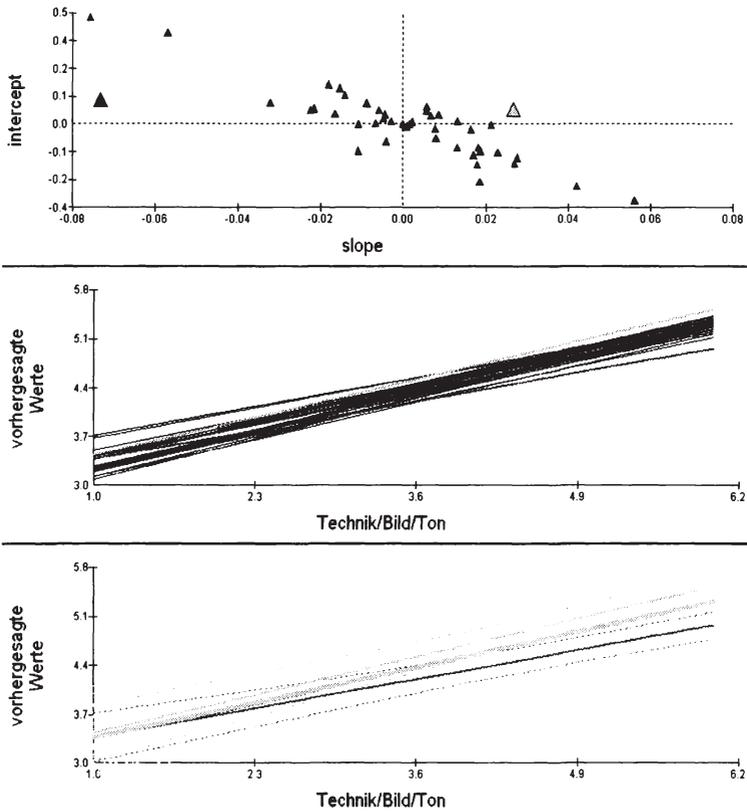


Diagramm 85: Standortvergleich bei gleichem Intercept

Es folgen die selbigen Schritte zum Vergleich von zwei weiteren Standorten in Diagramm 86 und abschließend Vergleichsinterpretationen zwischen den vier visualisierten Standorten. Es werden jene beiden Standorte gewählt, für welche der größte Unterschied in den *slope*- und *intercept*-Werten besteht. Auf das Vorliegen beider Extreme wird bereits durch den zuvor berechneten negativen Korrelationskoeffizienten hingewiesen. Der schwarz eingefärbte Standort weist den höchsten *intercept* und den niedrigsten *slope* auf. Der grau eingefärbte Standort weist den niedrigsten *intercept* und den höchsten *slope* auf. Dritte Graphik verdeutlicht den aus dem Korrelationskoeffizienten vorhergesagten Effekt. Je niedriger der *intercept*, desto höher der *slope*, und umgekehrt. In den besseren Bewertungen treffen die Regressionsgeraden aufeinander, in den schlechteren Bewertungen unterscheiden sich diese. Die dick markierte schwarze Linie gibt den allgemein erkennbaren Trend wieder.

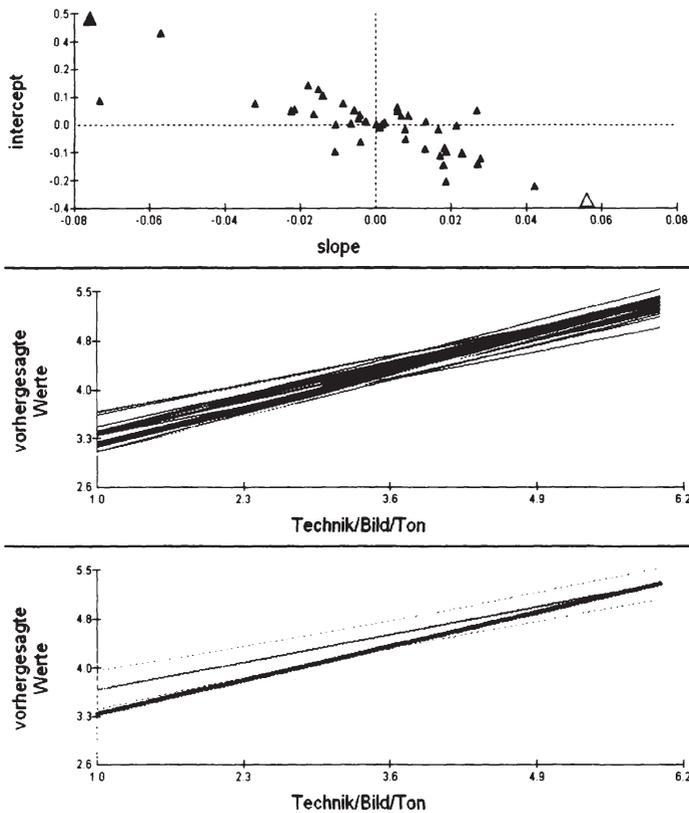


Diagramm 86: Standortvergleich bei gleichem Intercept und unterschiedlichen Slopes

Schlussfolgernd kann die Aussage gemacht werden, dass es in jenen Kinos, welche einen hohen *intercept*, aber einen niedrigen *slope* aufweisen, keine nennenswerte Auswirkung des Serviceleistungsattributes auf die Gesamtzufriedenheit gibt. Gleichgültig auf welchem Zufriedenheitslevel sich das Serviceleistungsattribut befindet, die Gesamtzufriedenheit bleibt annähernd konstant. Umgekehrt weisen jene Kinos, welche einen niedrigen *slope* und hohen *intercept* besitzen, eine Anfälligkeit auf Veränderungen des Serviceleistungsattributes Technik/Bild/Ton auf. Je höher die Zufriedenheit mit dem Serviceleistungsattribut, desto höher die Gesamtzufriedenheit. Je niedriger die Zufriedenheit mit dem Serviceleistungsattribut, desto niedriger die Gesamtzufriedenheit. Für letztgenannte Standorte ist es wichtig, auf diese Serviceleistungsattribute Acht zu geben und die erwarteten Levels zu erfüllen. Zuvor genannte besitzen größeren Spielraum. Ebenso ist hier aufgrund des

negativen linearen Zusammenhang zu bemerken, dass es kaum Kinos gibt, für welche sich die Zufriedenheitswerte die Waage halten. Entweder der *intercept* ist hoch und der *slope* niedrig, oder der *intercept* ist niedrig und der *slope* hoch. Dies bedeutet, dass es wie im ersten Fall, sehr wenig Einfluss gibt bzw. wie im zweiten Fall sehr hohen Einfluss gibt. Jene Kinos, mit niedrigen *intercepts* und *slopes* bzw. hohen *intercepts* und *slopes* fallen eher aus. Daraus resultiert die Empfehlung, dass jeweils das unter Betracht stehende Kino einem der beiden Fälle zuzuweisen ist, da die Handlungsempfehlungen in zwei konträre Richtungen zielen. Entweder ist äußerste Vorsicht geboten, was das Serviceleistungsattribut betrifft oder es kann als eher bedeutungslos, da zufriedengestellt, angesehen werden. Mittlere Aufmerksamkeit dem Serviceleistungsaspekt gegenüber gibt es nur in Ausnahmefällen, wo die Zufriedenheit schlecht ist, aber keine Auswirkungen hat oder ohnehin bereits sehr zufrieden erfüllt wird.

4.3.4.4. Externe Parameter und Interaktionseffekte - Kinodaten

Zur Erklärung der Gesamtzufriedenheit werden in Modell 6 zwei externe Items miteinbezogen. Erstes betrifft die Frage ‚Besitzen Sie die Kundenkarte eines Kinos?‘ mit den Antwortmöglichkeiten ‚Ja (1)‘ und ‚Nein (2)‘, zweites die Frage ‚Besuchen Sie im Allgemeinen...‘ mit den Antwortmöglichkeiten ‚...nur ein Stammkino/Lieblingskino? (1)‘, ‚...ein Stammkino/Lieblingskino und weitere verschiedene Kinos? (2)‘ und ‚...unterschiedliche Kinos? (3)‘. Die Referenzkategorien sind ‚keine Kundenkarte‘ und ‚Besuch von Stammkinos/Lieblingskinos und weiteren verschiedenen Kinos‘. Die geschätzten Parameter besitzen jeweils zwei Indizes.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons + \beta_{1j} tbt_{ij} + 0,058(0,022) karte_1_{ij} + 0,064(0,028) besuch_1_{ij} + -0,032(0,025) besuch_3_{ij}$$

$$\beta_{0ij} = 2,901(0,085) + u_{0ij}$$

$$\beta_{1ij} = 0,400(0,016) + u_{1ij}$$

$$\begin{bmatrix} u_{0ij} \\ u_{1ij} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,047(0,041) & \\ -0,008(0,008) & 0,002(0,002) \end{bmatrix}$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [1,394(0,016)]$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 47557,190(14986 \text{ of } 14986 \text{ cases in use})$$

Modell 6: Kundenkartenbesitz und Kinobesuchverhalten

Jene Personen, welche eine Kundenkarte besitzen, haben eine durchschnittlich um 0,058 Zufriedenheitspunkte höhere Gesamtzufriedenheit. Besuchen diese überwiegend ein Stammkino/Lieblingskino, ist deren Gesamtzufriedenheit um durchschnittlich 0,064 Zufriedenheitspunkte höher, besuchen diese überwiegend

unterschiedliche Kinos, ist deren Gesamtzufriedenheit um durchschnittlich 0,032 Zufriedenheitspunkte niedriger. Personen, welche eine Kundenkarte besitzen und überwiegend ein Stammkino/Lieblingskino besuchen, liegen um $0,058+0,064=0,122$ Gesamtzufriedenheitspunkte höher. Ein Kino, für welches die Gesamtzufriedenheit höher liegt, wird auch öfters besucht und wird zum Stammkino/Lieblingskino. Einflussrichtungsbedingt betrachtet führt umgekehrt der häufigere Besuch des gleichen Kinos zu höherer Zufriedenheit.

Das aktuelle Modell mit den Einflüssen der drei externen Kategorien besitzt einen *likelihood*-Wert von 47.557,190. Modell 7, ohne die drei neuen Parameter der zwei externen Variablen, führt zu einem Wert von 47.574,200. Die Reduktion von 17,01 bei drei Freiheitsgraden ergibt laut χ^2 -Verteilung eine signifikante Verbesserung mit einem Signifikanzwert von 0,00070340. Es ist erkennbar, dass die zwischen den Kinos liegende *intercept*-Wert-Variabilität gesunken ist, nämlich von 0,054 auf 0,047. Diese wird nun zusätzlich durch die in das Modell aufgenommenen beiden externen Items erklärt. Die Kundenkarte und die Kinowahl beeinflussen folglich den *intercept*.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons + \beta_{1ij} tbt_{ij}$$

$$\beta_{0ij} = 2,918(0,086) + u_{0ij} + e_{0ij}$$

$$\beta_{1ij} = 0,399(0,017) + u_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,054(0,044) & \\ -0,009(0,008) & 0,002(0,002) \end{bmatrix}$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [1,394(0,016)]$$

$$-2 * \text{loglikelihood(IGLS Deviance)} = 47574,200(14986 \text{ of } 14986 \text{ cases in use})$$

Modell 7: Einfaches Modell

Es stellt sich die Frage, ob dieser Einfluss auf den sechs Zufriedenheitslevels des Attributes Technik/Bild/Ton unterschiedlich ausfällt. Modell 8 schätzt Kinobesuchsverhalten und Kundenkartenbesitz in Interaktion mit der Technik/Bild/Ton-Zufriedenheit.

Der *likelihood*-Wert sinkt von 47.557,190 auf 47.551,790 um 5,4 bei 3 weiteren Freiheitsgraden. Aus der χ^2 -Verteilung ergibt sich eine Signifikanz von 0,14474. Die Modellveränderung ist nicht signifikant. Es ist jedoch erkennbar, dass die Kundenkarte, verglichen mit den Besuchertypen, in Bezug auf die Interaktionen die besseren Werte zeigt. Deshalb wird Modell 9 mit der alleinigen Interaktion zwischen Kundenkarte und Technik/Bild/Ton-Zufriedenheit geschätzt.

$$\begin{aligned}
ges_{ij} &\sim N(XB, \Omega) \\
ges_{ij} &= \beta_{0ij}cons + \beta_{1j}tbt_{ij} + -0,196(0,136)karte_1_{ij} + -0,119(0,177)besuch_1_{ij} + \\
&-0,200(0,152)besuch_3_{ij} + 0,047(0,025)karte_1.tbt_{ij} + 0,033(0,032)besuch_1.tbt_{ij} + \\
&0,031(0,028)besuch_3.tbt_{ij} \\
\beta_{0ij} &= 3,057(0,110) + u_{0j} + e_{0ij} \\
\beta_{1j} &= 0,371(0,021) + u_{1j} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,055(0,045) & \\ -0,009(0,008) & 0,002(0,002) \end{bmatrix} \\
[e_{0ij}] &\sim N(0, \Omega_e) : \Omega_e = [1,394(0,016)] \\
-2 * \text{loglikelihood (IGLS Deviance)} &= 47551,790 (14986 \text{ of } 14986 \text{ cases in use})
\end{aligned}$$

Modell 8: Interaktion mit Kinobesuchsverhalten und Kundenkartenbesitz

$$\begin{aligned}
ges_{ij} &\sim N(XB, \Omega) \\
ges_{ij} &= \beta_{0ij}cons + \beta_{1j}tbt_{ij} + -0,191(0,135)karte_1_{ij} + 0,063(0,028)besuch_1_{ij} + \\
&-0,031(0,025)besuch_3_{ij} + 0,046(0,025)karte_1.tbt_{ij} \\
\beta_{0ij} &= 2,983(0,096) + u_{0j} + e_{0ij} \\
\beta_{1j} &= 0,385(0,018) + u_{1j} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,054(0,0404) & \\ -0,009(0,008) & 0,002(0,002) \end{bmatrix} \\
[e_{0ij}] &\sim N(0, \Omega_e) : \Omega_e = [1,394(0,016)] \\
-2 * \text{loglikelihood (IGLS Deviance)} &= 47553,710 (14986 \text{ of } 14986 \text{ cases in use})
\end{aligned}$$

Modell 9: Interaktion mit Kundenkartenbesitz

Das Modell weist einen *likelihood*-Wert von 47.553,710 auf, verglichen mit 47.557,190 im Modell ohne Interaktion. Eine Devianz von 3,48 bei einem Freiheitsgrad kommt mit einem Signifikanzwert von 0,062115 laut χ^2 -Verteilung in eine signifikante Nähe. Die Ergebnisse der unterschiedlichen *slopes* für die einzelnen Gruppen werden trotzdem interpretiert. Der *slope* von Besuchern, welche keine Kundenkarte besitzen beträgt 0,385. Für jene Personen, welche eine Kundenkarte besitzen, ist dieser um 0,046 steiler und beträgt $0,385+0,046=0,413$. Der Interaktionsterm erklärt einen Teil der Streuung, welche zuvor noch in der zwischen den Kinos liegenden Variabilität gelegen ist. Der Einflusswert auf Standortebene ist nicht nennenswert gesunken und verbleibt bei 0,002. Diese Schlussfolgerungen betreffen rein den *slope*. Die vorhergesagte Regressionsgerade setzt sich nun für Besucher ohne Kundenkarte aus den geschätzten Werten $\hat{\beta}_0.cons + \hat{\beta}_1.tbt_{ij}$, die Regressionsgerade für Besucher mit Kundenkarte aus $\hat{\beta}_0.cons + \hat{\beta}_1.tbt_{ij} + \hat{\beta}_2.karte_1_{ij} + \hat{\beta}_3.karte_1.tbt_{ij}$ zusammen.

Diagramm 87 veranschaulicht den Effekt des Besitzes vs. Nichtbesitzes einer Kundenkarte zur Technik/Bild/Ton-Zufriedenheit.

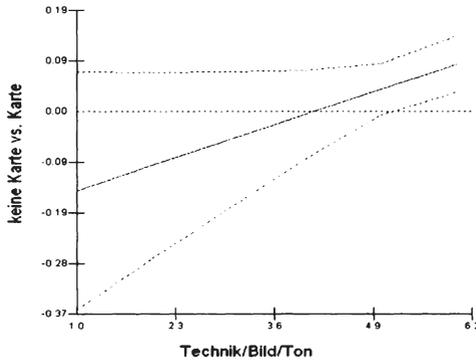


Diagramm 87: Kundenkarteneinfluss - Technik/Bild/Ton

Auf der y-Achse werden die prognostizierten Werte der Effekte aufgetragen, $\hat{\beta}_2 \text{ karte_1}_{ij} + \hat{\beta}_3 \text{ karte_1.tb1}_{ij}$. Auf der x-Achse die Technik/Bild/Ton-Zufriedenheiten. Die horizontale Gerade durch den Nullpunkt stellt die Gerade für die Nicht-Kundenkartenbesitzer dar. Der Effekt des Besitzes einer Kundenkarte verändert sich, je nachdem auf welchem Technik/Bild/Ton-Zufriedenheitslevel sich eine Person befindet. Jene Personen, welche eine Kundenkarte besitzen und sich auf dem höchsten Technik/Bild/Ton-Zufriedenheitsniveau befinden, liegen um etwa 0,09 Zufriedenheitspunkte in ihrer Gesamtzufriedenheit höher, verglichen mit jenen, welche keine Kundenkarte besitzen. Fällt die Zufriedenheit mit Technik/Bild/Ton unter einen Wert von etwa vier, ist deren Gesamtzufriedenheit besser, wenn sie keine Kundenkarte besitzen. Allerdings ist zu beachten, dass in den niedrigeren Technik/Bild/Ton-Zufriedenheitskategorien das strichliert eingezeichnete 95%-ige Konfidenzintervall deutlich zunimmt. Dies kann daraus resultieren, dass Personen, welche eine Kundenkarte besitzen, auch zufrieden mit Technik/Bild/Ton sind. Gerade deshalb besitzen diese eine Kundenkarte. Sinkt die Zufriedenheit mit Technik/Bild/Ton unter zuvor erwähnten Level, schwinden die positiven Effekte der Kundenkarte, da diese an Wert verliert und das Kino der Wahl eher zu Enttäuschung führt. Deshalb ist die Gesamtzufriedenheit auf diesem Level besser ohne Kundenkarte.

4.3.4.5. Varianzvisualisierungen der unterschiedlichen Ebenen - Kinodaten

Nun wird die Varianz zwischen den Standorten anhand der Technik/Bild/Ton-Zufriedenheit genauer untersucht. Die Varianzen auf Levelebene werden durch die Parameter der erklärenden Variablen beschrieben. Die Varianz auf Level 1

ergibt sich aus $\text{var}(e_{0ij}|\text{cons}) = \sigma_{e0}^2 \text{cons}^2$, jene auf Level 2 aus $\text{var}(u_{0j}|\text{cons} + u_{1j}|\text{bt}_{ij}) = \sigma_{u0}^2 \text{cons}^2 + 2\sigma_{u01} \text{cons} * \text{bt}_{ij} + \sigma_{u1}^2 \text{bt}_{ij}^2$. Die Konstante besteht aus lauter Einsen. Die Varianz zwischen den Standorten ist eine quadratische Funktion der Technik/Bild/Ton-Variablen, mit Koeffizienten der Level-2-Zufallsparameter. Der erste Term, auf der rechten Seite ist der *intercept* dieser Funktion, der zweite der lineare Term und der dritte der quadratische. Diagramm 88 visualisiert die Varianz auf Level-2-Ebene. Dies ist die zwischen den Standorten auffindbare Varianz, nach Berechnung dieser für die einzelnen Zufriedenheitswerte mit dem Serviceleistungsattribut Technik/Bild/Ton. Es ist erkennbar, dass bei niedrigen Technik/Bild/Ton-Zufriedenheitswerten, die Varianz zwischen den Standorten hoch ist. Dies bedeutet, dass die niedrigeren Zufriedenheitslevels die Standorte untereinander besser ausdifferenzieren, verglichen mit besseren Bewertungen. Der Standorteffekt ist für Besucher mit niedrigeren Zufriedenheitsbewertungen größer, verglichen mit Besuchern, die positivere Zufriedenheitsempfindungen erleben.

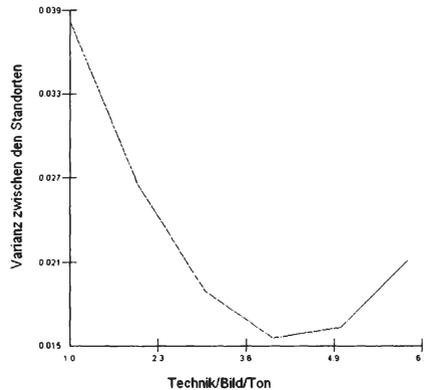


Diagramm 88: Level-2-Varianzveränderung - Technik/Bild/Ton

Bis dato wurde angenommen, dass die Abweichungen auf Level 1, also auf Besucherebene, konstant sind. Im Modell 10 wird überprüft, ob standortbezogen unterschiedliche Abweichungen existieren. Diese können sich beispielsweise bei tendenziell höheren bzw. niedrigeren Zufriedenheitswerten in einem Standort ändern. Dazu wird für den Technik/Bild/Ton-Koeffizienten, zusätzlich zum Level-2-Standortterm, ein Level-1-Besucherterm mitgeschätzt. Im Modell wird dies durch den zusätzlichen Term e_{ij} für den Koeffizienten β_{ij} zum Standortlevelterm u_{0j} repräsentiert. Die Varianz von u_{0j} , nämlich σ_{u1}^2 , stellt die Varianz der standortspezifischen *slopes* dar. Ebenso ist sie der Koeffizient in der quadratischen Funktion, welcher die Veränderung der standortspezifischen Variation im Hinblick auf die Technik/Bild/Ton-Zufriedenheit beschreibt. σ_{e1}^2

stellt die Veränderung zwischen den Besuchern in Bezug auf die Technik/ Bild/Ton-Zufriedenheit dar.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} \text{cons} + \beta_{1ij} \text{tbl}_{ij} + \beta_2 \text{karte_1}_{ij} + \beta_3 \text{besuch_2}_{ij} + \beta_4 \text{besuch_3}_{ij} + \beta_5 \text{karte_1.tbl}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = \beta_1 + u_{1j} + e_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma_{e0}^2 & \\ & \sigma_{e1}^2 \end{bmatrix}$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 47511,700 (14986 \text{ of } 14986 \text{ cases in use})$$

Modell 10: Level-1-Abweichungen

Die Varianzfunktion $\text{var}(e_{0ij} \text{cons} + e_{1j} \text{tbl}_{ij}) = \sigma_{e0}^2 \text{cons}^2 + 2\sigma_{e01} \text{cons} * \text{tbl}_{ij} + \sigma_{e1}^2 \text{tbl}_{ij}^2$ von Modell 10 wird visualisiert. Die untere Kurve zeigt die Varianz zwischen den Kinostandorten, welche auch im oberen Plot dargestellt wurde. Aufgrund der deutlich höheren Varianz zwischen den Besuchern ist die Krümmung aufgrund der Skalierung nicht mehr gut ersichtlich. Der Varianzgrößenvergleich der folgenden Graphik kann rechnerisch ermittelt werden.

$$\frac{S \text{ tan dort var ianz}}{S \text{ tan dort var ianz} + \text{Besucher var ianz}} \quad \text{vs.} \quad \frac{\text{Besucher var ianz}}{S \text{ tan dort var ianz} + \text{Besucher var ianz}}$$

In Diagramm 89 ist erkennbar, dass mit steigender Technik/Bild/Ton-Zufriedenheit die Varianz zwischen den Besuchern sinkt. Die höhere Varianz bleibt wiederum in den weniger zufriedenen Antwortkategorien. Dies war auch aus den vorangegangenen Modellen ersichtlich.

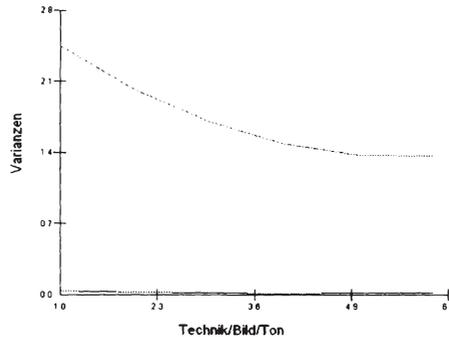


Diagramm 89: Level-1- und -2-Varianzveränderung - Technik/Bild/Ton

In Modell 11 werden die Einflüsse der Kundenkarte und der Technik/Bild/Ton-Zufriedenheit auf Level-1-Besucherebene analysiert.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + \beta_{1ij}tbt_{ij} + \beta_{2i}karte_1_{ij} + \beta_3besuch_1_{ij} + \beta_4besuch_3_{ij} + \beta_3karte_1.tbt_{ij} \\
 \beta_{0ij} &= \beta_1 + u_{0j} + e_{0ij} \\
 \beta_{1ij} &= \beta_2 + u_{1j} + e_{1ij} \\
 \beta_{2i} &= \beta_2 + e_{2ij}
 \end{aligned}$$

$$\begin{aligned}
 \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma_{e0}^2 & & \\ \sigma_{e01} & \sigma_{e1}^2 & \\ \sigma_{e02} & \sigma_{e12} & \sigma_{e2}^2 \end{bmatrix}
 \end{aligned}$$

-2 * loglikelihood (IGLS Deviance) = 47508,930 (14986 of 14986 cases in use)

Modell 11: Kundenkarteneinfluss auf Besucherebene - Technik/Bild/Ton

Negative Kovarianzen sind eigentlich nicht möglich (Rasbash et al., 2005), werden in Modell 12 aber zugelassen. Die Level-1-Varianzen müssen deshalb immer als Gesamtes mit den weiteren Parametern der Varianzfunktion betrachtet werden.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + \beta_{1ij}tbt_{ij} + \beta_{2i}karte_1_{ij} + 0,065(0,028)besuch_1_{ij} + \\
 &-0,030(0,25)besuch_3_{ij} + 0,041(0,026)karte_1.tbt_{ij} \\
 \beta_{0ij} &= 2,955(0,097) + u_{0j} + e_{0ij} \\
 \beta_{1ij} &= 0,390(0,018) + u_{1j} \\
 \beta_{2i} &= -0,160(0,142) + e_{2ij}
 \end{aligned}$$

$$\begin{aligned}
 \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,042(0,041) & \\ -0,007(0,007) & 0,002(0,001) \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 1,927(0,041) & & \\ -0,048(0,010) & 0 & \\ 0 & 0 & -0,062(0,035) \end{bmatrix}
 \end{aligned}$$

-2 * loglikelihood (IGLS Deviance) = 47523,570 (14986 of 14986 cases in use)

Modell 12: Kundenkarteneinfluss auf Besucherebene - Technik/Bild/Ton Ergebnis

Die Varianzfunktion lautet $\text{var}(e_{0ij}cons + e_{1ij}tbt_{ij} + e_{2ij}karte_1_{ij}) = \sigma_{e0}^2cons^2 + 2\sigma_{e01}cons * tbt_{ij} + \sigma_{e1}^2karte_1_{ij}^2$ und ergibt eine Besuchervarianz von 1,927-0,048*Technik/Bild/Ton-Wert für jene Personen, welche keine Kundenkarte besitzen. Die Kurve für die zwischen den Besuchern auftretende Varianz liegt für jene Personen, welche eine Kundenkarte besitzen niedriger.

Modell 13 dient der Überprüfung, ob die Gruppen auch unterschiedliche slopes aufweisen. Dafür wird ein komplexeres Modell für die Level-1-Varianz geschätzt.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij}cons + \beta_{1ij}tbt_{ij} + \beta_{2i}karte_1_{ij} + 0,065(0,028)besuch_1_{ij} + -0,030(0,025)besuch_3_{ij} + 0,040(0,026)karte_1.tbt_{ij}$$

$$\beta_{0ij} = 2,955(0,097) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = 0,390(0,018) + u_{1j}$$

$$\beta_{2i} = -0,160(0,142) + e_{2ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,042(0,042) & \\ -0,007(0,008) & 0,002(0,001) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 1,931(0,142) & & \\ -0,048(0,013) & 0 & \\ 0 & 0,001(0,022) & -0,076(0,0248) \end{bmatrix}$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 47523,560 (14986 \text{ of } 14986 \text{ cases in use})$$

Modell 13: Modell 12 mit Slopevariation

Die Varianzfunktion $\text{var}(e_{0ij}cons + e_{1ij}tbt_{ij} + e_{2ij}karte_1_{ij}) = \sigma_{e0}^2 cons^2 + 2\sigma_{e01} cons * tbt_{ij} + 2\sigma_{e12} tbt_{ij} * karte_1_{ij} + \sigma_{e2}^2 karte_1_{ij}^2$ mit der Level-1-Varianz der Kundenkarten-Nichtbesitzer beträgt $1,931 + 2 * (-0,048) * \text{Technik/Bild/Ton-Wert} = 1,931 - 0,096 * \text{Technik/Bild/Ton-Wert}$. Die Level-1-Varianz der Kundenkartenbesitzer liegt bei $1,931 + (2 * (-0,048) + 2 * 0,001) * \text{Technik/Bild/Ton-Wert} - 0,076 = 1,855 - 0,094 * \text{Technik/Bild/Ton-Wert}$. Beide zeigen eine negative Entwicklung in Diagramm 90.

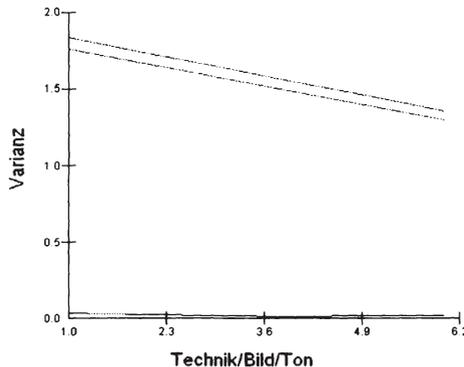


Diagramm 90: Varianzunterschiede Kundenkartenbesitzer vs. -nichtbesitzer

Die Varianzen beider Gruppen sinken mit steigender Zufriedenheit. Dies bedeutet, dass sie in den höheren Zufriedenheitslagen untereinander eine

geringere Ausdifferenzierung in Bezug auf die Gesamtzufriedenheit besitzen. Die Kundenkarten-Nichtbesitzer haben einen etwas, jedoch nicht merklich stärkeren *slope*. Dies bedeutet, dass diese in den höheren Zufriedenheitslagen untereinander eine geringere Variabilität in Bezug auf die Gesamtzufriedenheit besitzen. Da der Wert jedoch derartig gering ist, kann dieser Effekt hier nur zum besseren Verständnis, nicht jedoch zum Erkenntnisgewinn genutzt werden. Die etwas tiefer liegende Gerade der Kundenkartenbesitzer deutet darauf hin, dass deren Einfluss auf die Gesamtzufriedenheit, aus der Perspektive zwischen den einzelnen Besuchern dieser Gruppe heraus betrachtet, einander ähnlicher ist, verglichen mit den Kundenkarten-Nichtbesitzern. Kundenkartenbesitzer sind somit in sich homogener als die heterogeneren Kundenkarten-Nichtbesitzer.

Die zuvor entfernten *slope*-Teile der Kovarianzmatrix werden in Modell 14 hinzugefügt und in Diagramm 91 visualisiert. Aufgrund der *slope*-Werte zeigt sich keine lineare Veränderung der beiden Gruppen mehr.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij} cons + \beta_{1ij} tbt_{ij} + \beta_{2i} karte_1_{ij} + 0,067(0,028) besuch_1_{ij} + \\
 &-0,031(0,025) besuch_3_{ij} + 0,036(0,027) karte_1.tbt_{ij} \\
 \beta_{0ij} &= 2,935(0,097) + u_{0ij} + e_{0ij} \\
 \beta_{1ij} &= 0,394(0,018) + u_{1ij} + e_{1ij} \\
 \beta_{2i} &= -0,134(0,148) + e_{2ij} \\
 \begin{bmatrix} u_{0ij} \\ u_{1ij} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,038(0,041) & \\ -0,006(0,007) & 0,001(0,001) \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 2,966(0,373) & & \\ -0,288(0,075) & 0,052(0,015) & \\ 0,001(0,132) & -0,006(0,024) & 0,000(0,000) \end{bmatrix} \\
 -2 * \text{loglikelihood(IGLS Deviance)} &= 47508,920 (14986 \text{ of } 14986 \text{ cases in use})
 \end{aligned}$$

Modell 14: Kundenkartenbesitzeinfluss mit Slopevariation - Ergebnisse

Im Vergleich zum vorherigen Modell weisen Kundenkartenbesitzer als auch - nichtbesitzer in den schlechteren Technik/Bild/Ton-Zufriedenheitsbereichen eine ähnlich hohe Varianz auf. Diese driftet mit zunehmender Zufriedenheit auseinander und fällt für Kundenkartenbesitzer stärker ab. Es macht somit Sinn, die zusätzlichen Parameter für die *slope*-Werte mitzuschätzen. Werden nur die Gesamtzufriedenheitsmittelwerte der Kundenkartenbesitzer und -nichtbesitzer verglichen, ist erkennbar, dass erstere mit 5,05, verglichen mit letzteren mit 4,98, die höhere Gesamtzufriedenheit aufweisen. Die Standardabweichung beträgt 1,23. Im Datensatz befinden sich 10.839 Kundenkartennichtbesitzer und 4.147 - besitzerbewertungen.

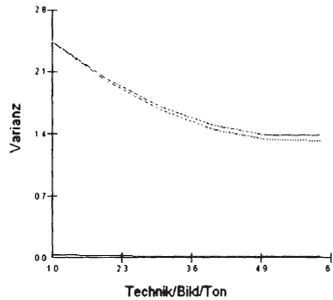


Diagramm 91: Kundenkartenbesitzer vs. -nichtbesitzer mit Slopevariation

$$\frac{5,05 - 4,98}{\sqrt{1,23^2 \left(\frac{1}{10.839} + \frac{1}{4.147} \right)}} = \frac{0,07}{0,02245885} = 3,11681141$$

Der Variationsvergleichswert beider Gruppen wird mit dem zum p -Wert zugehörigen Wert der Standardnormalverteilung verglichen. Je höher desto signifikanter. Der p -Wert beträgt 0,00091409, ist hoch signifikant und hat für den zweiseitigen Test die doppelte Größe, da *a priori* nicht spezifiziert wurde, ob Kundenkartenbesitzer oder -nichtbesitzer zufriedener bzw. unzufriedener sind. Das 95%-ige Konfidenzintervall für die Mittelwertdifferenzen in der Gesamtpopulation beträgt $0,07 \pm 1,96 * 0,02245885 = 0,025980654 / 0,044019346$. Ebenso kann das Modell als Regression formuliert werden, wobei die Eigenschaft des Kundenkartenbesitzes die Gesamtzufriedenheit beeinflusst. Dazu wird das *ordinary least squares*-Verfahren bzw. für in diesem Falle zwar kategoriale Daten, eine Varianzanalyse (ANOVA) durchgeführt. Über diese Modellformulierung könnten in Modell 15 zusätzlich weitere Gruppen und erklärende Variablen in das Modell aufgenommen werden. Die Referenzkategorie ist ‚keine Kundenkarte‘.

$$y_i = 4,985(0,012) + 0,065(0,022)x_{i1} + e_i$$

$$e_i \sim N(0, \sigma_e^2) \sigma_e^2 = 1,512(0,017)$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 48724,260 (14986 \text{ of } 14986 \text{ cases in use})$$

Modell 15: Kundenkartenvariation

Residuen sind normalverteilt mit Mittelwert null und Varianz σ^2 . Die beiden geschätzten Parameter β_0 und β_1 in der Regressionsgleichung beschreiben den fixen Teil des Modells, die Residuen und deren Varianz den zufälligen Teil. Der Mittelwert für die Kundenkartennichtbesitzer, der *intercept*, beträgt 4,985 und ist im Vergleich zu den Kundenkartenbesitzern um 0,065 Zufriedenheitspunkte niedriger. Dies entspricht der zuvor gerundeten Differenz von 0,07. Ebenso ist

die Residuenvarianz von 1,512 gleich groß, verglichen mit der zuvor gemessenen Standardabweichung der gesamten Stichprobe von 1,23². Um zu bestimmen, ob die Mittelwertabweichung beider Gruppen auch für die Grundgesamtheit gilt, wird hier der geschätzte Parameter 0,065 durch dessen Standardabweichung 0,022 dividiert. Das Ergebnis von 2,95454 liegt in der Nähe des Normalverteilungswertes.

Nun wird darauf aufbauend ein Modell mit mehr als zwei Ebenen gerechnet. Insgesamt gibt es 53 Kinostandorte. Die Mittelwerte der einzelnen Standorte verteilen sich wie im Diagramm 92 veranschaulicht im hohen Zufriedenheitsbereich. Neun wurden nicht besucht und es stehen keine Bewertungen zur Verfügung, was durch den linken Balken dargestellt wird.

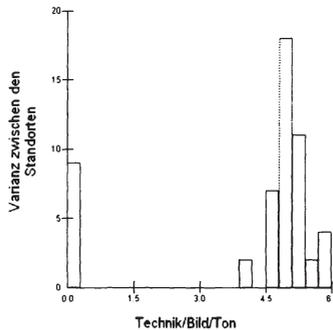


Diagramm 92: Mittelwertverteilung - Kinostandorte

4.3.4.6. Modellaufbau mit drei Ebenen - Kinodaten

In den nun folgenden Modellen wird, aufgrund unterschiedlicher im Modell enthaltener Variablen und dem späteren Aufbau eines Modells mit drei Ebenen, eine kleinere Probandengruppe aufgenommen. Um trotzdem die Vergleichbarkeit mit einfacheren Modellen zu ermöglichen, werden die Modellergebnisse des geschälerten Datensatzes teilweise nochmals angeführt. Die Probandenzahl wird durch die Anzahl jener Personen bestimmt, welche für alle Variablen im Endmodell vollständig beobachtete Werte enthalten. Damit die Modelle untereinander vergleichbar sind, dürfen die Probandenzahlen zwischen den Modellen nicht variieren. Dies ermöglicht die Vergleichbarkeit mit Hilfe der LR-*statistic*. Es werden lediglich jene 23 Standorte miteinbezogen, für welche es über hundert Besucherbewertungen gibt. Daraus werden jene Probanden verwendet, welche alle in den folgenden Modellen enthaltenen Fragen beantwortet haben. Die Standortmittelwerte können auch mit Hilfe des *single-level (fixed effects)* Regressionsmodells geschätzt werden. Die Dummy-

Variablen enthalten den Wert eins wenn das jeweilige Kino unter Betrachtung steht, ansonsten den Wert null. Der *intercept* gibt in diesem Fall den Wert für das Referenzkino Nr. 23 an. So können die Kinostandortindizes der geschätzten Standortparameter für die Interpretation übernommen werden. Um den Mittelwert eines bestimmten Kinostandortes zu erhalten, wird zum *intercept* der geschätzte Parameter für das jeweilige Kino addiert. Eine weitere Möglichkeit wäre, das Modell ohne *intercept* zu schätzen, um sofort die Mittelwerte der jeweiligen Standorte als Parameterschätzungen zu erhalten. Um jedoch einen besseren Vergleichsüberblick über die Lage des Standortes zu bekommen, wird Modell 16 mit *intercept* angeschrieben. Ist der geschätzte Parameter für einen Standort positiv (negativ), so liegt der Zufriedenheitsmittelwert dieses Standortes über (unter) dem Gesamtmittelwert aller Standorte.

$$\begin{aligned}
 ges_i = & 5,290(0,091) + -0,595(0,184)kino_1_i + -0,314(0,095)kino_2_i + -0,268(0,106)kino_3_i + \\
 & -0,240(0,152)kino_4_i + -0,558(0,163)kino_5_i + -0,586(0,190)kino_6_i + -0,282(0,096)kino_7_i + \\
 & -0,361(0,094)kino_8_i + -0,297(0,102)kino_9_i + -0,161(0,094)kino_10_i + -0,456(0,210)kino_11_i + \\
 & -0,379(0,121)kino_12_i + -0,322(0,157)kino_13_i + 0,036(0,159)kino_14_i + -0,273(0,128)kino_15_i + \\
 & -0,320(0,105)kino_16_i + -0,210(0,103)kino_17_i + -0,004(0,226)kino_18_i + -0,291(0,094)kino_19_i + \\
 & -0,614(0,107)kino_20_i + 0,165(0,206)kino_21_i + -0,286(0,105)kino_22_i
 \end{aligned}$$

$$e_i \sim N(0, \sigma_e^2) \sigma_e^2 = 1,506(0,017)$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 48127,580(14986 \text{ of } 14986 \text{ cases in use})$$

Modell 16: Mittelwertschätzung der Standortvariation

Folgend wird als Nullhypothese angenommen, dass es keine Abweichungen zwischen den jeweiligen Standorten gibt. Der *F*-Test verwendet die Summe der Quadrate der Abweichungen zwischen den Standorten und die Summe der Quadrate der Abweichungen zwischen den Besuchern innerhalb der Standorte. Aus Tabelle 47, welche die Ergebnisse der Varianzanalyse zeigt, sind jeweils die Mittelwerte der einzelnen Standorte und die Anzahl der Besucher abzulesen.

Um den *p*-Wert für die Unterschiede zwischen den Standorten zu erhalten, muss der Wert von 4,95 bei 22 und 14.820 Freiheitsgraden abgeglichen werden. Der Wert ist mit $1,9949e^{-013}$ hoch signifikant, was auf unterschiedliche Mittelwerte der Standorte hindeutet. Das mittlere Quadrat innerhalb der Standorte liegt bei 1,5082. Dieser ist, verglichen mit dem Fehlerwert aus Modell 16 unterschiedlich, da die ANOVA *restricted* IGLS (RIGLS) benutzt und Modell 16 IGLS.

	Df	SS	MS	F
Between groups	22	164,32	7,4689	4,95
Within groups	14798	22319	1,5082	
Total	14820	22483	1,5171	

Pooled within-group S.D. = 1,2281

Between-group variance component = 0,0099329

	N	Mean	S.E.M.
1	59	4,6949	
2	1690	4,9757	
3	502	5,0219	
4	101	5,0495	
5	82	4,7317	
6	54	4,7037	
7	1394	5,0079	
8	2602	4,9281	
9	701	4,9929	
10	2369	5,1283	
11	42	4,8333	
12	236	4,9110	
13	92	4,9674	
14	89	5,3258	
15	182	5,0165	
16	551	4,9691	
17	639	5,0798	
18	35	5,2857	
19	2162	4,9991	
20	472	4,6758	
21	44	5,4545	
22	540	5,0037	
23	183	5,2896	

General mean = 5,0005

Tabelle 47: Standortmittelwerte und Abweichungen

Zusätzlich werden die ineinander verschachtelten Modelle 16 und 17 über die χ^2 -basierte *LR-statistic* verglichen. Die *likelihood*-Differenz von 48.127,59-48.236,30=108,71 bei 22 Freiheitsgraden ist hoch signifikant bei 0,00000 und weist auf deutliche Unterschiede in der Grundgesamtheit hin.

$$ges_i = 5,000(0,010) + e_i$$

$$e_i \sim N(0, \sigma_e^2) \sigma_e^2 = 1,517(0,018)$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 48236,300(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 17: Standortvariation der Mittelwerte

Würde eine weitere erklärende Variable in das *fixed effects*-Modell auf einer höheren Ebene als Standortebene eingeführt werden, ergäbe die Parameterschätzung null, da diese durch Linearkombination der restlichen 22 Parameter dargestellt werden kann. Um Schätzungen höherer Levels zu ermöglichen, wird das Modell hier schrittweise nochmals aufgebaut. Für die größere Anzahl an Gruppen wird ein *random effects*- bzw. *multilevel*-Modell verwendet. Es wird davon ausgegangen, dass die Gruppeneffekte zufällig und normalverteilt sind. Die Residuen werden in zwei Teile getrennt. Ein Teil basiert auf Level-1-Personenebene, der andere Teil auf Level-2-Kinostandortebene. Die Unterschiede zwischen den Standorten werden, verglichen mit dem *fixed effects*-Modell, in Modell 18 verschieden behandelt.

$$y_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Modell 18: Random Effects-Modell

Die Standorteffekte sind in obigem Modell mit u_{0j} bezeichnet und stellen Zufallsvariablen dar. β_0 stellt den Mittelwert der Grundgesamtheit dar. Somit entfällt die Referenzkategorie. u_{0j} zeigt die standortbezogene Abweichung von diesem Mittelwert. Der *intercept* ist als zufällig auf dem Standortlevel definiert. Somit kann der Mittelwert eines Standortes zufällig um den Gesamtmittelwert mit u_{0j} variieren. Die Abweichungen der Besucher innerhalb eines Standortes beträgt e_{ij} . Das *variance components*-Modell splittet die Residuenvarianz auf den hierarchischen Ebenen in die Varianz zwischen den Gruppen (σ_{u0}^2) und die Varianz zwischen den Individuen innerhalb einer Gruppe (σ_e^2) auf. Die Ähnlichkeit zwischen den Individuen innerhalb einer Gruppe wird mit Hilfe der *intra-class*-Korrelation aus Formel 17 berechnet. Sie steht im Vergleich zu den Ähnlichkeiten von Individuen anderer Gruppen und repräsentiert jenen Teil der gesamten Residuenvarianz, welcher den Unterschieden zwischen den einzelnen Gruppen zuzuschreiben ist, auch *variance partition coefficient* (VPC) genannt.

$$\frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

Formel 17: Intra-Class-Korrelation

Der Gesamtmittelwert der Technik/Bild/Ton-Zufriedenheitsbewertungen beträgt in Modell 19 4,997. Die Mittelwerte der Standorte weichen mit einer geschätzten Varianz von 0,014 ab. Dieser Wert liegt bei einer Standardabweichung von 0,006 im signifikanten Bereich, $0,014/0,006=2,333$. Die Bewertungen der Probanden weichen innerhalb der Standorte mit einem Wert von 1,509 (0,018) ab.

$$ges_{ij} = \beta_{0j} + e_{ij}$$

$$\beta_{0j} = 4,997(0,031) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \sigma_{u0}^2 = 0,014(0,006)$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,509(0,018)$$

$$-2 * \log \text{likelihood (IGLS Deviance)} = 48189,570(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 19: Random Effects-Modell - Ergebnisse

Mit Hilfe eines LR-Tests wird Modell 19 mit 20 verglichen, in welchem σ_{u0}^2 auf null beschränkt wird und nicht von standortabhängigen Unterschieden ausgegangen wird. Die *likelihood*-Differenz von $48.236,300 - 48.189,570 = 46,73$ bei einem Freiheitsgrad ist hoch signifikant $8,1472e^{-12}$. Dies spricht für die Unterschiede auf Standortebene (Level 2).

$$ges_{ij} = 5,000(0,010) + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,517(0,018)$$

$$-2 * \log \text{likelihood (IGLS Deviance)} = 48236,300(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 20: Modellvergleich

Der VPC, als Prozentanteil der durch die Standorte begründete Varianz an der Gesamtvarianz, liegt bei $0,014/(0,014+1,509)=0,00919 \approx 1\%$. Die standortbedingten Abweichungen erklären folglich nur einen sehr geringen Anteil der Streuung.

Im Modell 21 werden zusätzliche Effekte auf Standortebene mitgeschätzt. Diese geben an, ob die Zufriedenheit mit Cineplexx-Standorten höher liegt, als jene, von Nicht-Cineplexx-Standorten. Im zuvor angesprochenen ANOVA-Modell ist dies aufgrund von Linearkombinationsproblemen nicht möglich.

$$ges_{ij} = \beta_{0j} + 0,052(0,061) \text{cineplexx}_j + e_{ij}$$

$$\beta_{0j} = 4,970(0,044) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \sigma_{u0}^2 = 0,014(0,006)$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,509(0,018)$$

$$-2 * \log \text{likelihood} = 48188,840(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 21: Random Effects-Modell mit drei Ebenen

Nicht-Cineplexx-Standorte werden durchschnittlich mit einem Zufriedenheitswert von 4,970 bewertet, Cineplexx-Standorte um 0,052 besser. Die Variation zwischen den Kinos u_{0j} ist nicht merklich gesunken. Die Cineplexx-Eigenschaft übt keinen wesentlichen Einfluss auf die Unterschiede zwischen den Standorten aus. Mehrebenenmodelle bieten also die Möglichkeit, Variation zwischen Standorten zu messen und gleichzeitig Variationen mit Hilfe von Gruppenlevel-Kovariaten mit einzubeziehen, um die Variation zwischen den Standorten zu erklären.

4.3.4.7. Residuenbetrachtungen - Kinodaten

Um genauere Vergleiche zwischen den Standorten anstellen zu können, müssen die standortspezifischen Residuen betrachtet werden. Durch Subtraktion der vorhergesagten Werte von den beobachteten Werten, in diesem Falle dem Gesamtmittelwert, werden diese aus Modell 19 berechnet. In *multilevel*-Modellen wird das Level-2-Residuum, auch *shrunken residual* genannt und in Bezug auf einen Standort wie in Formel 19 berechnet.

$$\hat{u}_{0j} = \frac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_{\epsilon} / n_j} r_{*j}$$

Formel 18: Shrunken Residual

n_j ist die Anzahl der Probanden aus Standort j . Zur Berechnung der rohen Residuen wird der geschätzte vom beobachteten Wert abgezogen, $r_{ij} = y_{ij} - \hat{y}_{ij}$. Das rohe Residuum für den j -ten Standort ist der Mittelwert aller r_{ij} dieses Standortes und mit r_{*j} bezeichnet. Dieser Verkleinerungsfaktor, mit welchem das rohe Residuum multipliziert wird, liegt zwischen null und eins. Ist die Varianz auf Probandenebene innerhalb eines Standortes hoch, verglichen mit der zwischen den Standorten, ist der Multiplikator kleiner. Er ist ebenfalls kleiner, wenn die Anzahl an Probanden pro Standort gering ist. Dies kann folgendermaßen erklärt werden. Angenommen zu einem Standort gibt es keine Bewertungen, so ist der beste Vorhersagewert der Gesamtmittelwert $\hat{\beta}_0$. Dieser ist aufgrund seiner Schätzung jedoch ungenau. Da jedoch davon ausgegangen wird, dass dieser Standort Teil der gesamten Standortpopulation ist und diese Information auch verwendet werden soll, wird eine gewichtete Kombination aus dem geschätzten Standortmittelwert und dem geschätzten Populationsmittelwert verwendet. Deshalb muss der beobachtete Standortwert in Richtung des Zentrums der Population verkleinert werden. Die beste Art und Weise dafür stellt der erwähnte Verkürzungsfaktor dar. Somit entsteht der zuvor erwähnte Unterschied in der Berechnung zwischen ANOVA und *multilevel*-Modellen. Angenommen der Standort mit den höchsten Abweichungen wurde lediglich von zwei Probanden bewertet, würde eine ANOVA den Mittelwert dieser Stichprobe

berechnen und eine hohe Standardabweichung annehmen. Der *shrinkage factor* bewegt deshalb den Mittelwert in die Richtung des Gesamtmittelwertes aller Standorte. Die Level-1-Residuen ergeben sich aus $\hat{\epsilon}_{ij} = r_{ij} - \hat{u}_{0j}$. Die Berechnung von Residuen ist auf jedem Level möglich.

Zusätzlich kann die Standardabweichung dazu berechnet werden. Die *comparative standard deviation* (SD) der Residuen, $\hat{u}_{0j} - u_{0j}$, wird für Schlussfolgerungen bezüglich u_{0j} benutzt. Das standardisierte Residuum wird als $\hat{u}_{0j} / SD(\hat{u}_{0j})$ definiert und kann zur Kontrolle der Normalverteilung geplottet werden. In Diagramm 93 wird der Multiplikator der Standardabweichungen auf 1,96 gesetzt und die Level 2 Residuen mit 95%-igen Konfidenzintervallen in aufsteigender Ordnung eingezeichnet.

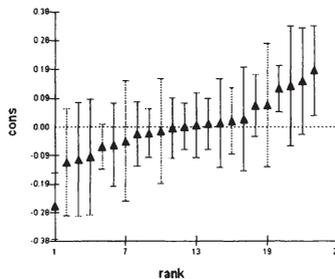


Diagramm 93: Level-zwei-Residuen und Konfidenzintervalle

Am linken Rand befindet sich ein Standort und am rechten Rand zwei Standorte, deren Konfidenzintervalle ihrer Residuen die strichlierte Null-Linie nicht überschneiden. Da diese Residuen die Standortabweichungen vom vorhergesagten fixen Parameter β_0 darstellen, bedeutet dies, dass diese signifikant vom Durchschnitt auf einem 5%-igen Level abweichen.

Es kann die Annahme überprüft werden, ob die geschätzten Residuen einer Normalverteilung folgen. Mit Hilfe eines Normalwahrscheinlichkeitsplots werden die ansteigenden Residuen gegen die zugehörigen Punkte auf der Normalverteilungskurve geplottet. Folgen die Residuen einer Normalverteilung, stellt diese Punktwolke eine Gerade dar. In Diagramm 94 finden sich in der linken Graphik die Level-1-Residuen, in der rechten Graphik die Level 2 Residuen. Normalverteilung kann nur schwer angenommen werden. Die Ordinalität der Skala kommt deutlich zum Ausdruck.

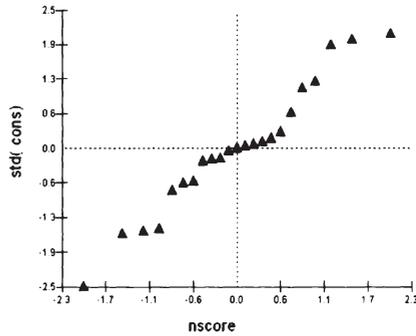
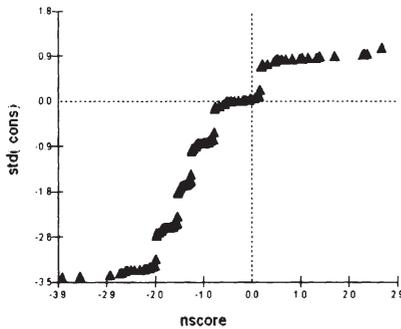


Diagramm 94: Residuen

Folgend wird nochmals ein Modell aufgebaut und die Technik/Bild/Ton-Zufriedenheit dazu verwendet, die Variable Gesamtzufriedenheit in Modell 22 zu erklären.

$$ges_i = 2,912(0,063) + 0,384(0,011)tb_i + e_i$$

$$e_i \sim N(0, \sigma_e^2) \sigma_e^2 = 1,410(0,016)$$

$$-2 * \log \text{likelihood} = 47147,280(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 22: Technik/Bild/Ton-Zufriedenheit und Gesamtzufriedenheit

Die Regressionsgerade setzt sich aus $ges = 2,912 + 0,384 * \text{Technik/Bild/Ton-Zufriedenheit}$ zusammen. Das bedeutet, dass mit jeder Technik/Bild/Ton-Zufriedenheitssteigerung von eins, auf einer Skala von ‚nicht zufrieden (1)‘ bis ‚sehr zufrieden (6)‘, die Gesamtzufriedenheit vom Wert 2,912 weg, um einen Wert von 0,384 multipliziert mit der Technik/Bild/Ton-Bewertung steigt. Die Werte der Gesamtzufriedenheit variieren mit einem Wert von 1,410 um den Gesamtdurchschnitt. Im Modell 23 wird der Standort als Level 2 miteinbezogen. Der *intercept* des *random effects*-Modell variiert zufällig über die Standorte.

$$ges_{ij} = \beta_{0j} + \beta_1 tb_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$ges_{ij} = \beta_{0j} + 0,392(0,012)tb_{ij} + e_{ij}$$

$$\beta_{0j} = 2,920(0,069) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \sigma_{u0}^2 = 0,014(0,006)$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,401(0,016)$$

$$-2 * \log \text{likelihood} = 47093,460(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 23: Technik/Bild/Ton-Zufriedenheit und Gesamtzufriedenheit mit Standortabweichungen

Jeder Standort hat einen eigenen *intercept*, β_j . Sie weisen einen einheitlichen hoch signifikanten *slope* von 0,392 (0,012) auf. Der *intercept*-Mittelwert aller Standorte beträgt 2,920 (0,069). Die vorhergesagte durchschnittliche Gerade aus

dem fixen *intercept* und *slope*, β_0 und β_1 , wird in Diagramm 96 erzeugt und gegen die Technik/Bild/Ton-Bewertungen abgeblendet.

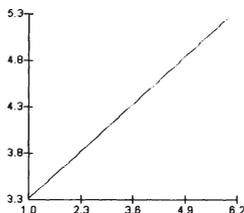


Diagramm 95: Variance Components Modell

Die einzelnen Standort-*intercepts* sind die Level-2-Residuen, u_{0j} , mit einer Varianz von 0,014 (0,006). Die Originalbewertungen variieren von diesen Geraden mit 1,401 (0,016), den Level-1-Residuals, e_{ij} . Vergleicht man den *likelihood*-Wert des linearen Regressionsmodells, 47.147,280, mit dem Wert des *multilevel*-Modells, 47.093,460, in welchem zusätzlich die Variation der *intercepts* auf Standortebene miteinbezogen wurde, erhält man mit einer Differenz von 53,82 bei einem Freiheitsgrad, σ^2_{u0} , einen hoch signifikanten Unterschied, $2,1972e^{-13}$. Der Standort stellt somit einen wesentlichen Unterschied, zusätzlich zur Miteinbeziehung der Zufriedenheit mit Technik/Bild/Ton, dar. Das eben erläuterte *variance components*-Modell, beschreibt eine einfache Variation auf Level-2-Ebene. Die Standortresiduen werden in Tabelle 48 ausgegeben. Der vorhergesagte Wert für Standort j beträgt $g\hat{e}s_{ij} = (2,920 + \hat{u}_{0j}) + 0,392 * Technik / Bild / Ton_Zufriedenheit_{ij}$.

Standort	1	2	3	4	5
Residuum	-0,045278	-6,287614e-02	4,439335e-02	3,564484e-02	-3,575953e-02
Standort	6	7	8	9	10
Residuum	-6,069775e-02	-4,706495e-02	-0,1500398	-0,1202492	-2,736012e-03
Standort	11	12	13	14	15
Residuum	1,338171e-02	-2,652187e-02	5,157511e-02	0,1438896	3,548896e-02
Standort	16	17	18	19	20
Residuum	-3,609444e-02	2,024222e-02	9,131587e-02	-9,535619e-02	-0,2484571
Standort	21	22	23		
Residuum	0,1136535	-6,553777e-02	0,2065998		

Tabelle 48: Standortresiduen

$g\hat{e}s_{ij} = (2,920 - 0,045278) + 0,392 * Technik / Bild / Ton_Zufriedenheit_{ij}$ stellt die vorhergesagte Gerade für Standort eins dar. In Diagramm 96 werden alle Standortlinien, mit der über alle Standorte fixen Schätzung und der geschätzten standortspezifischen

Abweichung $ges_{ij} = \hat{\beta}_{0j} + \hat{\beta}_1 tbt_{ij}$, visualisiert. Es ergeben sich 23 parallele Standortlinien. Linke Graphik zeigt die vorhergesagten Punkte der Standorte aller sechs Zufriedenheitsbewertungen. In der rechten Graphik sind die Punkte für die standortspezifischen Berechnungen für sämtliche Zufriedenheitsbewertungen verbunden.

Im Modell 24 wird die Variable ‚Cineplexx-Kino vs. Nicht-Cineplexx-Kino‘ miteinbezogen, welche zusätzliche Abweichungen erklären soll. Es ist erkennbar, dass die Miteinbeziehung keinen signifikanten Unterschied macht.

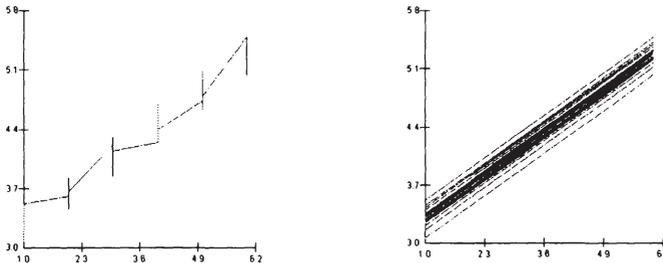


Diagramm 96: Vorhergesagte Standortwerte für die Zufriedenheitsbewertungen

$$ges_{ij} = \beta_{0j} + 0,392(0,012)tbt_{ij} + 0,008(0,061)cineplexx_j + e_{ij}$$

$$\beta_{0j} = 2,916(0,075) + u_{0j}$$

$$u_{0j} \sim N(0, \sigma^2_{u0}) \sigma^2_{u0} = 0,014(0,006)$$

$$e_{ij} \sim N(0, \sigma^2_e) \sigma^2_e = 1,401(0,016)$$

$$-2 * \log likelihood = 47093,560(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 24: Einfluss der Cineplexxeigenschaft

Betrachtet man nur ein Level im Modell 25, so sind die geschätzten Parameter sehr ähnlich. Wird der Cineplexx-Parameter im Modell 24, 0,008 (0,061), mit jenem aus Modell 25, 0,011 (0,021), verglichen, zeigt sich, dass in letzterem, in welchem nur ein Level berücksichtigt wird, die Standardabweichung deutlich niedriger ist. Der Einflusswert der Cineplexx-Eigenschaft im *single level*-Modell würde deutlich überschätzt werden. Die OLS-Regression ist folglich nicht geeignet, erklärende Variable auf Level 2 miteinzubeziehen. Generell ist auch anzumerken, dass im *single level*-Modell die fixen Koeffizienten auf Level 1 unterschätzt werden.

$$ges_{ij} = 2,906(0,064) + 0,384(0,011)tbt_{ij} + 0,011(0,021)cineplexx_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma^2_e) \sigma^2_e = 1,409(0,016)$$

$$-2 * \log likelihood = 47146,980(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 25: Ein Level

Im Folgenden wird für den geschmälernten Datensatz überprüft, ob es standortspezifisch unterschiedliche *slopes* gibt und folglich die Koeffizienten der Technik/Bild/Ton-Zufriedenheit standortabhängig variieren. Dies wird mit Hilfe eines *fixed effects*-Modells errechnet, in welchem es Dummy-Variablen für alle Standorte gibt und ebenso viele Interaktionen zwischen der Technik/Bild/Ton-Zufriedenheit und den Dummyvariablen. Somit resultieren für 23 Standorte 46 zu schätzende fixe Effekte. Hier wird jedoch davon ausgegangen, dass die Standorte eine Auswahl aus einer großen Population von Standorten ist. Im *fixed effects*-Modell gibt es keine Zusammenhänge zwischen den Standorten und es wird nicht berücksichtigt, dass diese derselben Population entspringen. Um dies zu vermeiden und weiters auch den Einfluss von standortklärenden Variablen zu ermöglichen, wird ein *multilevel*-Modell gebaut.

Geht man im Modell 26 davon aus, dass den zufällig gewählten Standorten eine Gesamtpopulation zugrunde liegt, muss die Technik/Bild/Ton-Zufriedenheit auch zufällig auf Level 2 sein, wodurch β_1 einen Index j erhält.

$$ges_{ij} = \beta_{0j} + \beta_{1j}tb_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ & \sigma_{u1}^2 \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$-2 * \log \text{likelihood} = 47146,980(14821 \text{ of } 14821 \text{ cases in use})$$

$$ges_{ij} = \beta_{0j} + \beta_{1j}tb_{ij} + e_{ij}$$

$$\beta_{0j} = 2,886(0,084) + u_{0j}$$

$$\beta_{1j} = 0,399(,016) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,041(0,038) & \\ -0,007(0,007) & 0,002(0,001) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,400(0,016)$$

$$-2 * \log \text{likelihood} = 47091,430(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 26: Technik/Bild/Ton-Zufriedenheit zufällig auf Level zwei

Der *slope* mit 0,399 ist ähnlich dem zuvor errechneten Wert von 0,392 aus dem Modell mit lediglich *intercept*-standortvariierenden Werten und fixem *slope*. Die Standort-*slopes* variieren um diesen Mittelwert mit einer Varianz von 0,002 (0,001). Die Standort-*intercepts* variieren mit einer Varianz von 0,041 (0,038). Die Kovarianz ist negativ mit -0,007 (0,007). Somit weisen Standorte mit generell höherer Zufriedenheit niedrigere Zufriedenheitsveränderungen auf,

wenn die Zufriedenheit mit Technik/Bild/Ton variiert bzw. in Standorten mit generell niedrigerer Zufriedenheit ist die Steigung der Gesamtzufriedenheit bei Verbesserung der Technik/Bild/Ton-Zufriedenheit höher. Es ergibt sich eine sehr niedrige Korrelation zwischen *intercept* und *slope* von $-0,007 / \sqrt{0,002 * 0,041} = -0,00006338769$. Die Variation der standortspezifischen *slopes* wird in der Visualisierung sehr gering ausfallen. Die Probanden variieren um ihre Standortgerade mit einer Varianz von 1,400 (0,016). Der *likelihood*-Modellvergleich mit nur einem *slope*, 47.093,46, und dem aktuellen Modell, 47.091,43, ergibt eine Differenz von 2,03. Es sind zwei weitere Parameter, nämlich die Varianz der *slope*-Residuen, u_{1j} , und die Kovarianz mit den *intercept*-Residuen, u_{0j} , in das Modell eingeflossen. Die *likelihood*- bzw. Devianzveränderung ist χ^2 -verteilt mit zwei Freiheitsgraden und stellt die Nullhypothese auf, dass die zusätzlichen Parameter über die Population null sind. Die Veränderung ist mit 0,36240 nicht signifikant. Somit wird das komplexere Modell zugunsten des einfacheren verworfen. Die vorhergesagten Werte $\hat{g}^s_{ij} = \hat{\beta}_{0j} + \hat{\beta}_{1j}bt_{ij}$ finden sich in Diagramm 97.

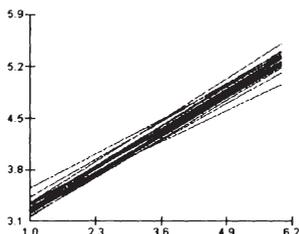


Diagramm 97: Vorhergesagte Werte für das Modell 26

Die Abweichungen bezüglich der Standorte halten sich, wie schon durch die Modellvergleiche gezeigt, in Grenzen. Die erste Graphik in Diagramm 98 stellt die Ergebnisse eines *random intercept fixed slope*-Modells dar. Jene Standorte mit der höchsten Zufriedenheitsabweichung nach oben bzw. unten werden mit einem großen Dreieck markiert. Die durchschnittliche Gerade ist in der unteren Graphik durch die Gerade etwa in der Mitte aller Geraden gekennzeichnet.

4.3.4.8. *Intercept-slope*-Zusammenhang - Kinodaten

Diagramm 99 zeigt ein *random slope random intercept*-Modell. Der Residuenplot, obere Graphik, beinhaltet zwei Level-2-Residuen, nämlich *intercepts* und *slopes* für die unterschiedlichen Standorte. Wiederum werden die zwei Standorte markiert. Vergleicht man die beiden Standorte mit den großen Dreiecken, so zeigt sich, dass beide bei niedriger Technik/Bild/Ton-Zufriedenheit, die gleiche Gesamtzufriedenheit aufweisen. In höheren Technik/

Bild/Ton-Zufriedenheitsbereichen driften die beiden Geraden auseinander. Auf den rechts markierten Standort hat die Technik/Bild/Ton-Zufriedenheit einen größeren Einfluss als auf den links markierten.

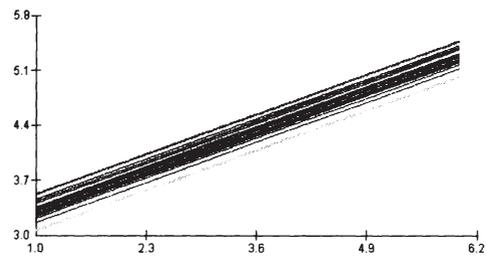
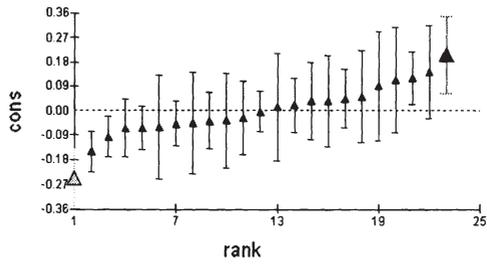


Diagramm 98: Random Intercept Fixed Slope

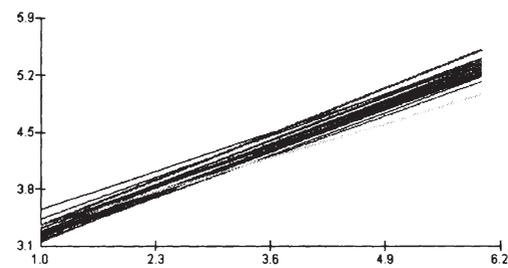
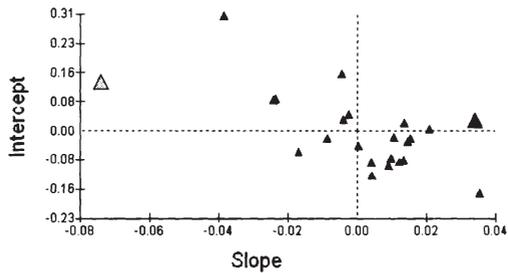


Diagramm 99: Random Slope Random Intercept

Quadrant I: Probanden sind mit jenen Kinos, welche sich im Quadranten links oben befinden überdurchschnittlich zufrieden. Die Technik/Bild/Ton-Zufriedenheit spielt dabei aber eine unterdurchschnittlich wichtige Rolle. Eine Steigerung der Technik/Bild/Ton-Qualität führt zu keinem starken Anstieg der Gesamtzufriedenheit. Der Kinostandort wird andere Möglichkeiten finden müssen, sofern er anstrebt, die Gesamtzufriedenheit zu verbessern.

Quadrant II: Probanden sind mit jenen Kinos, welche sich im Quadranten rechts oben befinden, überdurchschnittlich zufrieden. Zusätzlich spielt die Zufriedenheit mit Technik/Bild/Ton eine überdurchschnittlich große Rolle. Eine Steigerung der Technik/Bild/Ton-Qualität führt zusätzlich zu höherer Gesamtzufriedenheit. Der Kinostandort kann durch Verbesserungen noch höhere Gesamtzufriedenheit erzielen.

Quadrant III: Probanden sind mit jenen Kinos, welche sich im Quadranten links unten befinden unterdurchschnittlich zufrieden. Die Technik/Bild/Ton-Zufriedenheit hat nur unterdurchschnittlichen Einfluss auf die Gesamtzufriedenheit. Somit müssen sich diese Standorte andere Möglichkeiten suchen, um durch höhere Qualität bessere Gesamtzufriedenheit zu erzielen, sofern sie ihre ohnehin niedrigere Gesamtzufriedenheit verbessern wollen.

Quadrant IV: Probanden sind mit jenen Kinos, welche sich im Quadranten rechts unten befinden unterdurchschnittlich zufrieden. Die Technik/Bild/Ton-Zufriedenheit kann jedoch eine überdurchschnittlich höhere Gesamtzufriedenheit bewirken. Eine Steigerung der Technik/Bild/Ton-Qualität führt somit auch zu höherer Gesamtzufriedenheit. Dies kann ein Kinostandort sinnvoll nutzen.

In Hinblick auf die *intercept*- und *slope*-Werte, werden hier die vier Quadranten in weitere Gruppen eingeteilt. Diagramm 100 visualisiert untenstehende Erläuterungen.

Gruppe I: Quadrant I und Quadrant IV beschreiben jene Standorte, welche ihre Technik/Bild/Ton-Zufriedenheit bereits ausgereizt haben bzw. auf dem Weg dahin sind.

Gruppe II: Quadrant II und Quadrant III beschreiben jene Kinos, in welchen die Technik/Bild/Ton-Zufriedenheit entweder eine Auswirkung auf die Gesamtzufriedenheit hat, oder nicht.

Gruppe III: Ist die Verteilung der Standorte allgemein eher horizontal angeordnet, befinden sich die Standorte auf einem ausgeglichenen Zufriedenheitslevel, wenn die Technik/Bild/Ton-Zufriedenheit niedrig ist. Steigt die Technik/Bild/Ton-Zufriedenheit jedoch, differenzieren sich die Standorte untereinander aus.

Gruppe IV: Ist die Verteilung der Standorte allgemein eher vertikal angeordnet, befinden sich die Standorte auf unterschiedlichen Gesamtzufriedenheitsebenen. Die Technik/Bild/Ton-Zufriedenheit spielt in der Ausdifferenzierung jedoch eine untergeordnete Rolle.

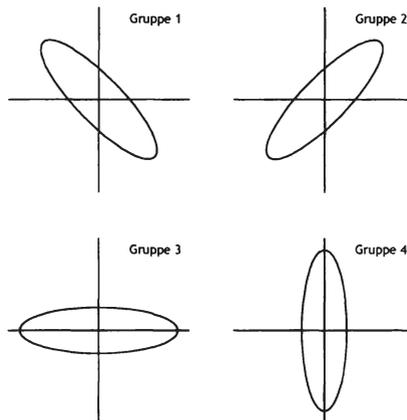


Diagramm 100: Quadranteneinteilungen für Intercept und Slope Kombinationen

Diagramm 101 zeigt die 95%-igen Konfidenzintervalle beider Standorte und die Gerade des Gesamtmittelwertes der *intercepts* und *slopes*. Erkennbar ist, dass sich beide Standorte in den niedrigeren Zufriedenheitsbereichen mit deren Konfidenzintervallen überschneiden, welche in diesen Bereichen größer sind. Sie tun dies jedoch nicht mehr in höheren Bereichen. Die beiden Standorte differenzieren sich in den höheren Zufriedenheitsbereichen folglich aus.

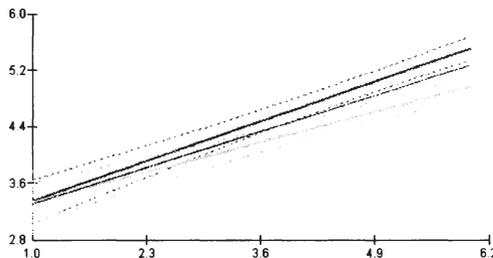


Diagramm 101: Konfidenzintervalle

Im Modell 27 ist auf die Frage ‚Besuchen Sie im Allgemeinen...?‘, die Eigenschaft ‚...ein Stammkino/Lieblingskino und weitere verschiedene Kinos?‘ die Referenzkategorie. Der Parameter ‚Stammkino‘ steht für die Antwort ‚...nur ein Stammkino/Lieblingskino‘, ‚Verschiedene‘ steht für ‚...unterschiedliche Kinos?‘ und befindet sich auf Probandenlevel mit Index ij . Der Parameter ‚Cineplexx‘ gibt an, ob das besuchte Kino ein Cineplexx-Kino ist, oder nicht. Er besitzt lediglich den Index j , da er auf Standortebene analysiert wird.

$$ges_{ij} = \beta_{0j} + \beta_{1j}bt_{ij} + 0,067(0,028)Stammkino_{ij} + -0,043(0,025)Verschiedene_j + 0,007(0,061)cineplexx_j + e_{ij}$$

$$\beta_{0j} = 2,883(0,089) + u_{0j}$$

$$\beta_{1j} = 0,400(0,016) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,037(0,037) & \\ -0,006(0,007) & 0,002(0,001) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma^2_e) \sigma^2_e = 1,399(0,016)$$

$$-2 * \log \text{likelihood} = 47080,460(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 27: Level-1- und -2-Einflüsse

Kinobesucher, welche unterschiedliche Kinos besuchen und zusätzlich ein Lieblingskino besitzen, bewerten ihre Gesamtzufriedenheit mit einem Cineplexx-Kino mit $2,883+0,007+0,4*Technik/Bild/Ton_Zufriedenheit$, mit einem Nicht-Cineplexx-Kino mit 0,007 Zufriedenheitspunkten schlechter. Stammkinobesucher, welche ein Cineplexx-Kino bewerten, sind mit vorhergesagten $2,883+0,067+0,007+0,4*Technik/Bild/Ton_Zufriedenheit$, die Zufriedensten. Stammkinobesucher, welche kein Cineplexx-Kino bewerten, weisen eine Zufriedenheit von $2,883+0,067+0,4*Technik/Bild/Ton_Zufriedenheit$ auf. Probanden, welche meistens verschiedene Kinos besuchen und ein Cineplexx-Kino bewerten, liegen in ihrer Zufriedenheit bei $2,883-0,043+0,007+0,4*Technik/Bild/Ton_Zufriedenheit$. Bewerten diese kein Cineplexx-Kino, liegt ihr Zufriedenheitswert bei $2,883-0,043+0,4*Technik/Bild/Ton_Zufriedenheit$. Vergleicht man dieses Modell mit dem Modell ohne Einfluss des Kinobesuchungsverhaltens und der Cineplexx-Eigenschaft, zeigt sich eine Devianz von $47.091,430-47.080,460=10,970$ bei 3 Freiheitsgraden und ist mit 0,011889 signifikant. Die Gesamtzufriedenheit wird unter Miteinbeziehung der beiden Eigenschaften besser erklärt. Die drei zusätzlichen Parameter verändern den *intercept*.

Es stellt sich nun die Frage, ob sich der Effekt entlang der Technik/Bild/Ton-Zufriedenheitsskala ändert. Im Modell 28 wird die Interaktion zwischen Technik/Bild/Ton-Zufriedenheit und der Cineplexx-Eigenschaft mitgeschätzt.

$$\begin{aligned}
ges_{ij} &= \beta_{0j} + \beta_{1j}tbt_{ij} + 0,099(0,165)cineplexx_j + 0,067(0,028)Stammkino_{ij} + \\
&- 0,043(0,025)Verschiedene_{ij} + -0,019(0,032)cineplexx.tbt_{ij} + e_{ij} \\
\beta_{0j} &= 2,8229(0,128) + u_{0j} \\
\beta_{1j} &= 0,411(0,025) + u_{1j} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,033(0,035) & \\ -0,005(0,006) & 0,001(0,001) \end{bmatrix} \\
e_{ij} &\sim N(0, \sigma_e^2) \sigma_e^2 = 1,399(0,016) \\
-2 * \log likelihood &= 47080,120(14821 \text{ of } 14821 \text{ cases in use})
\end{aligned}$$

Modell 28: Interaktion Technik/Bild/Ton-Zufriedenheit und Cineplexx-Eigenschaft

Die Devianz sinkt um 0,36 bei einem Freiheitsgrad und ist nicht signifikant. Die *slope*-Variabilität zwischen den Standorten ist jedoch von 0,002 auf 0,001 gesunken. Dieser Teil wird durch den zusätzlichen Interaktionsterm erklärt. Modell 29 misst die Interaktion zwischen Besuchsverhalten und Technik/Bild/Ton-Zufriedenheit. Die Devianz sinkt um 1,43 bei einem Freiheitsgrad und ist nicht signifikant.

$$\begin{aligned}
ges_{ij} &= \beta_{0j} + \beta_{1j}tbt_{ij} + 0,007(0,061)cineplexx_j + -0,112(0,177)Stammkino_{ij} + -0,169(0,155)Verschiedene_{ij} + \\
&0,033(0,032)Stammkino.tbt_{ij} + 0,023(0,028)Verschiedene.tbt_{ij} + e_{ij} \\
\beta_{0j} &= 2,942(0,103) + u_{0j} \\
\beta_{1j} &= 0,389(0,019) + u_{1j} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,038(0,037) & \\ -0,006(0,007) & 0,002(0,001) \end{bmatrix} \\
e_{ij} &\sim N(0, \sigma_e^2) \sigma_e^2 = 1,399(0,016) \\
-2 * \log likelihood &= 47079,030(14821 \text{ of } 14821 \text{ cases in use})
\end{aligned}$$

Modell 29: Interaktion Besucherverhalten und Technik/Bild/Ton-Zufriedenheit

Im Modell 30 wird die Interaktion des Kundenkarten-Besitzes mit der Technik/Bild/Ton-Zufriedenheit gemessen. Die Frage lautete, ‚Besitzen Sie die Kundenkarte eines Kinos?‘, mit Antwortmöglichkeiten ‚ja (1)‘ und ‚nein (2)‘.

Der *slope*-Wert für Probanden ohne Kundenkarte ist 0,387, für jene mit Kundenkarte ist dieser um 0,163 kleiner. Somit spielt die Technik/Bild/Ton-Zufriedenheitsveränderung bei Kunden mit Kundenkarte eine geringere Rolle, als bei jenen Personen, welche keine Kundenkarte besitzen. Die Gesamtzufriedenheit hat sich bei Kundenkartenbesitzern bereits stärker gefestigt und ist durch die Technik/Bild/Ton-Zufriedenheit schwerer zu irritieren. Zusätzlich hebt eine hohe Technik/Bild/Ton-Zufriedenheit die Gesamtzufriedenheit durch den Interaktionsterm bei Kundenkartenbesitzern an.

$$ges_{ij} = \beta_{0j} + \beta_{1j}tbt_{ij} + 0,005(0,060)cineplexx_j + 0,063(0,028)Stammkino_j + -0,036(0,025)Verschiedene_j + 0,042(0,025)karte.tbt_{ij} + -0,163(0,136)karte_{ij} + e_{ij}$$

$$\beta_{0j} = 2,935(0,100) + u_{0j}$$

$$\beta_{1j} = 0,387(0,018) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,040(0,038) & \\ -0,007(0,007) & 0,002(0,001) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,398(0,016)$$

$$-2 * \log likelihood = 47069,700(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 30: Interaktion Kundenkarte und Technik/Bild/Ton-Zufriedenheit

Der Einfluss der beiden zusätzlichen Terme, $\beta_2 cineplexx_j + \beta_3 cineplexx.tbt_{ij}$, wird ausgehend vom einfachen Modell ohne Stammkinoeigenschaft und Kartenbesitz, getrennt visualisiert. Dafür wird die Cineplexx-Technik/Bild/Ton-Interaktion im Modell 31 aufgenommen.

$$ges_{ij} = \beta_{0j} + \beta_{1j}tbt_{ij} + \beta_2 cineplexx_j + \beta_3 cineplexx.tbt_{ij} + e_{ij}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$-2 * \log likelihood = 47091,040(14821 \text{ of } 14821 \text{ cases in use})$$

$$ges_{ij} = \beta_{0j} + \beta_{1j}tbt_{ij} + 0,107(0,167)cineplexx_j + -0,019(0,032)cineplexx.tbt_{ij} + e_{ij}$$

$$\beta_{0j} = 2,822(0,129) + u_{0j}$$

$$\beta_{1j} = 0,410(0,025) + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,035(0,036) & \\ -0,006(0,007) & 0,002(0,001) \end{bmatrix}$$

$$e_{ij} \sim N(0, \sigma_e^2) \sigma_e^2 = 1,400(0,016)$$

$$-2 * \log likelihood = 47091,030(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 31: Interaktion Cineplexx-Eigenschaft und Technik/Bild/Ton-Zufriedenheit

In Diagramm 102 ist der Effekt der Cineplexx-Eigenschaft auf die Technik/Bild/Ton-Zufriedenheit mit seinem 95%-igen Konfidenzintervall visualisiert. Der Einfluss der Cineplexx-Eigenschaft auf die Technik/Bild/Ton-Zufriedenheit ist geringer, wenn die Person mit Technik/Bild/Ton sehr zufrieden ist. Ist die Person unzufrieden, so stehen, verglichen mit den Nicht-Cineplexx-Kinos, die Cineplexx-Kinos besser da. Trotz dessen, dass die Cineplexx-Bewertung mit Technik/Bild/Ton-Zufriedenheit niedrig ist, stehen diese in der Gesamt-

zufriedenheit besser da, als wenn ein anderer Standort betroffen wäre. Steigt in beiden Cineplexx-Eigenschaften die Zufriedenheit mit Technik/Bild/Ton, hat diese positive Veränderung einen besseren Einfluss auf Nicht-Cineplexx-Kinos, verglichen mit Cineplexx-Kinos. Somit holen Nicht-Cineplexx-Kinos bei steigender Zufriedenheit mit Technik/Bild/Ton ihren Gesamtzufriedenheitsrückstand auf.

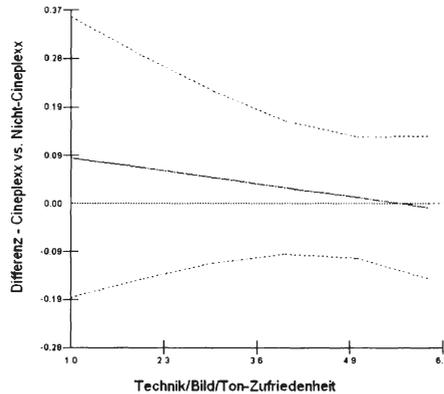


Diagramm 102: Einfluss Cineplexx-Eigenschaft

Bis dato wurde von einer einheitlichen Variation auf Level 1 ausgegangen. An dieser Stelle werden für die Eigenschaft, Cineplexx-Kino oder nicht, unterschiedliche Varianzen errechnet. Dies wurde bis dato durch den einheitlichen *intercept*, β_0 , und einem Probandenlevelresiduum, e_i , mit zugrunde liegender Varianz, $\sigma^2_{e_i}$, verhindert. Die Standardabweichung in beiden Gruppen ist jedoch unterschiedlich, wie aus Tabelle 49 entnommen werden kann.

	Gesamtzufriedenheit		
	0	1	TOTALS
Anzahl	4784	10037	14821
Mittelwert	4,98	5,01	5,00
Standardabweichung	1,26	1,22	1,23

Tabelle 49: Standardabweichungsvergleich

Das folgende Modell schätzt diese Werte, nämlich unterschiedliche *intercepts* für die beiden Gruppen. Dafür wird eine Dummyvariable für Cineplexx-Besucher und eine für Nicht-Cineplexx-Besucher erzeugt. Die Variable ‚Cineplexx‘ hat den Wert eins wenn es ein Cineplexx-Kino betrifft, und null, wenn dies nicht zutrifft. Die Variable ‚Nicht-Cineplexx‘ ist gegensätzlich codiert. Zusätzlich werden unterschiedliche Varianzen für beide Gruppen berechnet. Beide β

erhalten einen Index j . Die dritte Zeile gibt die Berechnung für Cineplexx-, die vierte für Nicht-Cineplexx-Mittelwerte.

Die Varianz der Cineplexx-Bewertungen ist $\text{var}(e_{0i}) = \sigma^2_{e0}$, die Varianz der Nicht-Cineplexx-Bewertungen ist $\text{var}(e_{1i}) = \sigma^2_{e1}$, und die Kovarianz auf Probandenebene zwischen dem Cineplexx-Residuum und dem Nicht-Cineplexx-Residuum ist σ_{e0e1} . Da es keine Kovarianz gibt, da eine Bewertung nicht gleichzeitig zu einem Cineplexx-Standort und zu einem Nicht-Cineplexx-Standort abgegeben werden kann, wird diese aus Modell 32 entfernt.

$$ges_i \sim N(XB, \Omega)$$

$$ges_i = \beta_{0i} \text{cineplexx}_i + \beta_{1i} \text{Nicht-Cineplexx}_i$$

$$\beta_{0i} = \beta_0 + e_{0i}$$

$$\beta_{1i} = \beta_1 + e_{1i}$$

$$\begin{bmatrix} e_{0i} \\ e_{1i} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} & \\ & \sigma^2_{e1} \end{bmatrix}$$

$$ges_i \sim N(XB, \Omega)$$

$$ges_i = \beta_{0i} \text{cineplexx}_i + \beta_{1i} \text{Nicht-Cineplexx}_i$$

$$\beta_{0i} = 5,011(0,012) + e_{0i}$$

$$\beta_{1i} = 4,978(0,018) + e_{1i}$$

$$\begin{bmatrix} e_{0i} \\ e_{1i} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 1,481(0,021) & \\ & 0 & 1,593(0,033) \end{bmatrix}$$

$$-2 * \log \text{likelihood (IGLS Deviance)} = 48225,280(14821 \text{ of } 14821 \text{ cases in use})$$

Modell 32: Entfernung der Kovarianz

Die geschätzten Werte stimmen mit den zuvor errechneten Mittelwerten überein. Auch deren Standardabweichungen weisen die gleichen Werte auf, $\sqrt{1,481} = 1,22$ und $\sqrt{1,593} = 1,26$. Die Varianz der Gesamtzufriedenheit ist $\text{var}(y_i) = \sigma^2_{e0} x_{0i} + \sigma^2_{e1} x_{1i}$. Im Modell 33 wird wiederum ein *random slope and intercept-Modell* spezifiziert.

$$y_{ij} \sim N(XB, \Omega)$$

$$y_{ij} = \beta_{0ij} x_0$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$[u_{0j}] \sim N(0, \Omega_u) : \Omega_u = [\sigma^2_{u0}]$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [\sigma^2_{e0}]$$

Modell 33: Random Slope und Intercept Modell

Das geschätzte Modell 34 beinhaltet die Technik/Bild/Ton-Zufriedenheit, welche zur Erklärung der Standortvariabilität verwendet wird.

$$\begin{aligned}
ges_{ij} &\sim N(XB, \Omega) \\
ges_{ij} &= \beta_{0ij} \text{cons} + \beta_{1ij} \text{ibt}_{ij} \\
\beta_{0ij} &= 2,886(0,084) + u_{0ij} + e_{0ij} \\
\beta_{1ij} &= 0,399(0,016) + u_{1ij} \\
\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,040(0,038) & \\ & -0,007(0,007) \quad 0,002(0,001) \end{bmatrix} \\
[e_{0ij}] &\sim N(0, \Omega_e) : \Omega_e = [1,400(0,016)] \\
-2 * \log \text{likelihood} (\text{IGLS Deviance}) &= 47091,430 \text{ (14821 of 14821 cases in use)}
\end{aligned}$$

Modell 34: Technik/Bild/Ton-Zufriedenheit zur Erklärung der Standortvariabilität

Die Varianz auf Probandenebene (Level 1) ist e , $\text{var}(e_{0ij} \text{cons}) = \sigma^2_{e_0} \text{cons}^2$, die Varianz auf Standortebene (Level 2) ist u , $\text{var}(u_{0j} \text{cons} + u_{1j} \text{ibt}_{ij}) = \sigma^2_{u_0} \text{cons}^2 + 2\sigma_{u_0 u_1} \text{cons} * \text{ibt}_{ij} + \sigma^2_{u_1} \text{ibt}_{ij}^2$. Letzterer Teil in Klammern ist die Summe der zwei Zufallskoeffizienten mit deren erklärenden Variablen multipliziert. *cons* ist ein allgemeiner *intercept*, ein Vektor mit durchgehendem Wert eins und die zwischen den Standorten bestehende Varianz ist eine quadratische Funktion der Technik/Bild/Ton-Zufriedenheiten mit Koeffizienten aus Level 2. Der Schnittpunkt der quadratischen Funktion ist $\sigma^2_{u_0}$, der lineare Term ist $2\sigma_{u_0 u_1}$, und der quadratische Term ist $\sigma^2_{u_1}$. Dies wird in Diagramm 103 visualisiert.

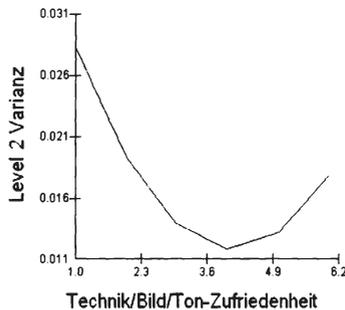


Diagramm 103: Level-2-Varianzveränderung mit Technik/Bild/Ton-Zufriedenheit

Je höher die Technik/Bild/Ton-Zufriedenheit, desto höher sind die Unterschiede zwischen den Standorten, bezogen auf die Gesamtzufriedenheit. Selbiges gilt für niedrige Zufriedenheitswerte. Der VPC, Standortvarianz/(Standortvarianz+ +Probandenvarianz), sinkt vorerst mit höheren Zufriedenheitswerten und steigt dann wieder an. Der Standorteffekt ist für Probanden mit höheren und niedrigeren Zufriedenheitsbewertungen stärker, bei mittleren Zufriedenheitswerten schwächer. Es ist auch möglich, dass die Level-1-Variation sich mit der Technik/Bild/Ton-Zufriedenheit ändert. Somit können die Probanden

Veränderungen über deren Standortgerade aufweisen. Dafür wird der Technik/Bild/Ton-Parameter zufällig auf Probandenebene in Modell 35 gesetzt und erhält den Index j . Nun besitzt β einen Zufallsterm auf Level-2-Ebene und einen auf Level-1-Ebene. Die Varianz von u_{1j} , nämlich σ^2_{u1} , ist einerseits die standortunterschiedliche *slope*-Variation und gleichzeitig der Koeffizient der quadratischen Funktion, welcher beschreibt, wie sich die zwischenstandort-spezifische Varianz mit der Technik/Bild/Ton-Zufriedenheit ändert. Erste Interpretation macht auf Level 1 keinen Sinn, da eine Person nicht variieren kann. Zweite Interpretation schon, nämlich wie sich die Level 1 Variation mit den erklärenden Variablen ändert.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} \text{cons} + \beta_{1ij} \text{tbt}_{ij}$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = \beta_1 + u_{1j} + e_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} & \\ & \sigma^2_{e1} \end{bmatrix}$$

$$-2 * \log \text{likelihood (IGLS Deviance)} = 47131,550 \text{ (14821 of 14821 cases in use)}$$

Modell 35: Zufälliger Technik/Bild/Ton-Parameter auf Probandenebene

Die Varianzfunktion $\text{var}(e_{0ij} \text{cons} + e_{1ij} \text{tbt}_{ij}) = \sigma^2_{e0} \text{cons}^2 + 2\sigma_{e01} \text{cons} * \text{tbt}_{ij} + \sigma^2_{e1} \text{tbt}_{ij}^2$ wird in Diagramm 104 gezeigt.

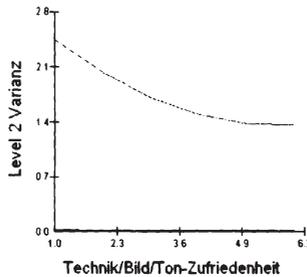


Diagramm 104: Level-1- und -2-Varianz über Technik/Bild/Ton-Zufriedenheit

Die untere kaum sichtbare Kurve zeigt wiederum die standortunterschiedliche Variation. Die obere Kurve stellt die Variation zwischen den Probanden dar. Sie ist aufgrund des quadratischen Wertes, σ^2_{e1} , von 0,052, gebogen. Die Variation zwischen den Probanden, innerhalb der Standorte, nimmt mit zunehmender Technik/Bild/Ton-Zufriedenheit ab. Dies flaut jedoch mit steigender

Zufriedenheit ab. Im Modell 36 wird die Probandenlevelvarianz als Funktion der Besuchereigenschaft, als auch der Technik/Bild/Ton-Zufriedenheit dargestellt.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij}cons + \beta_{1ij}tbt_{ij} + \beta_{2ij}besuch_{ij}$$

$$\beta_{0ij} = 2,979(0,090) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = 0,404(0,016) + u_{1j} + e_{1ij}$$

$$\beta_{2ij} = -0,056(0,016) + e_{2ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,027(0,036) & \\ -0,004(0,006) & 0,001(0,001) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 3,122(0,584) & & \\ -0,325(0,086) & 0,053(0,015) & \\ -0,012(0,126) & 0,016(0,018) & -0,019(0,034) \end{bmatrix}$$

$$-2 * \log likelihood (IGLS Deviance) = 47030,780 \text{ (14821 of 14821 cases in use)}$$

Modell 36: Probandenlevelvarianz als Funktion der Besuchereigenschaft und Technik/Bild/Ton-Zufriedenheit

Die Varianzfunktion lautet $\text{var}(e_{0ij}cons + e_{1ij}tbt_{ij} + e_{2ij}besuch_{ij}) = \sigma^2_{e0}cons^2 + 2\sigma_{e01}cons * tbt_{ij} + \sigma^2_{e1}tbt_{ij}^2 + 2\sigma_{e02}cons * besuch_{ij} + 2\sigma_{e12}tbt_{ij} * besuch_{ij} + \sigma^2_{e2}besuch_{ij}^2$. Daraus werden die Varianzen auf Probandenebene für die Stammkinobesucher, für jene die verschiedene Kinos besuchen und für jene welche verschiedene Kinos besuchen und zusätzlich ein Lieblingskino haben, berechnet. Diagramm 105 zeigt die Veränderung über die Technik/Bild/Ton-Zufriedenheitsskala hinweg.

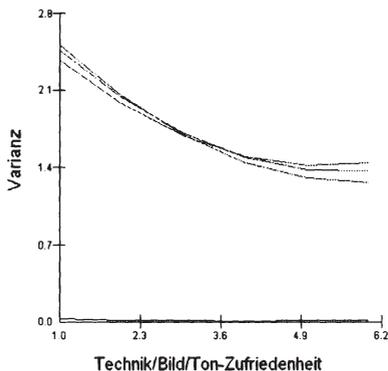


Diagramm 105: Varianzvisualisierung der Besuchertypen

In den mittleren Technik/Bild/Ton-Zufriedenheitsbereichen überschneiden sich die Kurven. In den Randbereichen hingegen finden sich sehr wohl Unterschiede. Dies bedeutet, dass die standortbedingte Probandenvarianz für alle drei Gruppen

in den mittleren Zufriedenheitsbereichen ähnlich ist. Steigt bzw. sinkt die Technik/Bild/Ton-Zufriedenheit, driften die drei Gruppen auseinander, und es ergeben sich die Standorte vergleichend differenziertere Probandengruppen. Modell 37 stellt dies für die Cineplexx-Eigenschaft dar.

$$ges_{ij} \sim N(\lambda B, \Omega)$$

$$ges_{ij} = \beta_{0ij}cons + \beta_{1ij}tbt_{ij} + \beta_{2i}cineplexx_j$$

$$\beta_{0ij} = 2,856(0,090) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = 0,403(0,016) + u_{1j} + e_{1ij}$$

$$\beta_{2i} = 0,011(0,061) + e_{2ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,027(0,036) & \\ -0,004(0,006) & 0,001(0,001) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 3,278(0,413) & & \\ -0,312(0,077) & 0,052(0,015) & \\ -0,234(0,135) & 0,036(0,024) & 0,000(0,000) \end{bmatrix}$$

$$-2 * \log \text{likelihood (IGLS Deviance)} = 47044,070 \text{ (14821 of 14821 cases in use)}$$

Modell 37: Probandenlevelvarianz als Funktion der Cineplexx-Eigenschaft und Technik/Bild/Ton-Zufriedenheit

Aus der Varianzformel $\text{var}(e_{0ij}cons + e_{1ij}tbt_{ij} + e_{2ij}cineplexx_j) = \sigma_{e0}^2 cons^2 + 2\sigma_{e01} cons * tbt_{ij} + \sigma_{e1}^2 tbt_{ij}^2 + 2\sigma_{e02} cons * cineplexx_j + 2\sigma_{e12} tbt_{ij} * cineplexx_j + \sigma_{e2}^2 cineplexx_j^2$ mit Werten $3,278 + 2 * (-0,312) * \text{Technik/Bild/Ton_Zufriedenheit} + 0,052 * \text{Technik/Bild/Ton_Zufriedenheit}^2 + 2 * (-0,234) * 1 + 2 * 0,036 * \text{Technik/Bild/Ton_Zufriedenheit} * 1 + +0,000 * 1^2$ wird die Kurve für das Cineplexx-Kino in Diagramm 106 visualisiert.

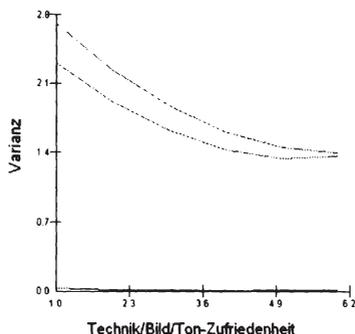


Diagramm 106: Varianzvisualisierung für unterschiedliche Kintypen

Für ein Nicht-Cineplexxkino fällt die Berechnung kürzer aus, da einige der Terme aufgrund der fehlenden Cineplexx-Eigenschaft wegfallen, $3,278 + 2 * (-0,312) * \text{Technik/Bild/Ton-Zufriedenheit} + 0,052 * \text{Technik/Bild/Ton-}$

Zufriedenheit². Die Kurve der Cineplex-Kinos liegt niedriger. Dies bedeutet, dass Cineplex-Kinos, die Auswirkung der Technik/Bild/Ton-Zufriedenheit betreffend, robuster und konstanter sind, verglichen mit Nicht-Cineplex-Kinos. Die Variabilität der Zufriedenheit unter Betrachtung der Cineplex-Eigenschaft ist in den niedrigen Technik/Bild/Ton-Zufriedenheitsbereichen sehr hoch. Diese nimmt jedoch mit zunehmender Technik/Bild/Ton-Zufriedenheit ab. Die Varianzen beider Gruppen, sowohl der Cineplex- als auch Nicht-Cineplex-Bewertungen sinken mit steigender Zufriedenheit. Die Nicht-Cineplex-Bewerter besitzen einen stärkeren *slope*. Dies bedeutet, dass diese auf dem Weg in die höheren Zufriedenheitslagen untereinander eine schneller fallende geringere Variabilität in Bezug auf die Gesamtzufriedenheit, entwickeln. Allgemein weisen die Gruppen in den höheren Zufriedenheitslagen untereinander eine geringere Ausdifferenzierung in Bezug auf die Gesamtzufriedenheit auf. Würde man eine derartige Level-1-Heterogenität nicht miteinbeziehen, würde dies möglicherweise zu einer Erhöhung der Level-2-Varianzen führen, da die Heterogenität zwischen den Probanden in die Heterogenität zwischen den Standorten einfließt. Dies führt zu ungewollten Verzerrungen.

4.3.4.9. Behandlung ordinaler Skalen - Kinodaten

Da im Kinodatensatz nicht von kontinuierlich skalierten Variablen ausgegangen werden kann, werden nun Modelle mit ordinal skalierten Daten vorgestellt. Die Häufigkeiten der Antworten auf die Gesamtzufriedenheit findet sich in Diagramm 107.

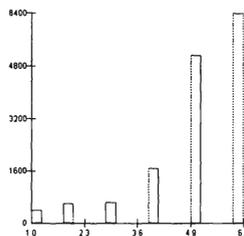


Diagramm 107: Antwortverteilung

4.3.4.10. Quadratische und kubische Terme - Kinodaten

Es liegt keine Normalverteilung vor. Somit folgt erst eine Transformierung der Gesamtzufriedenheit auf normalverteilte Werte. Der Mittelwert der Technik/Bild/Ton-Zufriedenheit wird berechnet und die Variable anschließend normalisiert. Dafür wird die Inverse der kumulierten Standardnormalverteilung berechnet, also normalverteilte Häufigkeiten. Dies sind die Wahrscheinlichkeiten (Probit) einen Wert im Intervall 0 bis 1 zu erhalten. Um nicht-lineare

Beziehungen zu ermöglichen, werden in Modell 38 neue Variablen mit dem Quadrat und Kubik berechnet.

$$\begin{aligned}
 gesnormal_i &\sim N(X\beta, \Omega) \\
 gesnormal_i &= \beta_0 \cdot cons + 0,324(0,013)tbnormal_i + 0,058(0,043)tb^2 + 0,027(0,017)tb^3, \\
 \beta_{0i} &= -0,047(0,018) + e_{0i} \\
 [e_{0i}] &\sim N(0, \Omega_e) : \Omega_e = [0,652(0,008)] \\
 -2 * \log likelihood(IGLS Deviance) &= 35720,920(14821 \text{ of } 14821 \text{ cases in use})
 \end{aligned}$$

Modell 38: Quadratische und kubische Terme

Es zeigt sich das Ergebnis eines Modells mit nur einer Ebene mit normalisierten Daten. Der quadratische und kubische Term sind nicht signifikant. Der Einfluss von *tbnormal*, die Technik/Bild/Ton-Zufriedenheit, ist hoch signifikant.

4.3.4.11. Ordered multinomial model - Kinodaten

Dieses vereinfachte Modell wird nun als ordinal-skaliertes Modell aufgebaut. Die zu inkludierenden Variablen Gesamtzufriedenheit und Technik/Bild/Ton-Zufriedenheit sind jeweils auf einer ‚nicht zufrieden (1)‘ bis ‚sehr zufrieden (6)‘, Skala definiert, $s = 1, \dots, t$. Die letzte Kategorie, in diesem Falle t ist die Referenzkategorie. Die Wahrscheinlichkeit dass Proband i in die Kategorie s

fällt ist $E(y_i) = \gamma_i^{(s)} = \sum_{h=1}^s \pi_i^{(h)}$, $s = 1, \dots, t-1$. $y_i^{(s)}$ sind die beobachteten

kumulativen Wahrscheinlichkeiten für den i -ten Probanden. Die Wahrscheinlichkeiten der Kategorien als kumulierte Wahrscheinlichkeiten sind $\pi_i^{(h)} = \gamma_i^{(h)} - \gamma_i^{(h-1)}$; $1 < h < t$; $\pi_i^{(1)} = \gamma_i^{(1)}$; $\gamma_i^{(t)} = 1$, das *proportional odds*-Modell mit *logit link* $\gamma_i^{(s)} = \{1 + \exp[-\alpha^{(s)} + (X\beta)_i]\}^{-1}$ oder *logit* $(\gamma_i^{(s)}) = \alpha^{(s)} + (X\beta)_i$, Weiter wird eine Multinomialverteilung der Kategorienwahrscheinlichkeiten unterstellt. Die kumulierten Wahrscheinlichkeiten haben eine Kovarianzmatrix von $cov(y_i^{(s)}, y_i^{(r)}) = \gamma_i^{(s)}(1 - \gamma_i^{(r)})/n_i$; $s \leq r$. Jede Person befindet sich nun auf Level-2-Ebene und die Antwortmöglichkeiten mit den Responsekategorien stellen Level 1 dar. Dafür werden neue Variablen erstellt, welche eine Länge von $14.821 \times 5 = 74.105$ besitzen, für jede Antwortkategorie eine Datenzeile. $\logit(\gamma_{s_j})$ ist der *logit* der erwarteten Wahrscheinlichkeit, dass ein Proband j die Zufriedenheit mit fünf oder schlechter bewertet. Im Modell 39 gibt es nur eine Antwortmöglichkeit auf die Frage nach den Kategorien, so wird für n_j die Konstante mit den Werten eins eingefügt.

Im Modell 39, auch *ordered proportional odds*-Modell genannt, wird die Auswahl getroffen, einen separaten *intercept* für jede der fünf Antwortkategorien

zu berechnen. Es kann jedoch auch ein *intercept* über alle gebildet werden. Die Schätzwerte des Modells 39 befinden sich in Modell 40.

$$resp_{ij} \sim \text{Ordered Multinomial}(cons_j, \pi_{ij})$$

$$\gamma_{1j} = \pi_{ij}; \gamma_{2j} = \pi_{1j} + \pi_{2j}; \gamma_{3j} = \pi_{1j} + \pi_{2j} + \pi_{3j}; \gamma_{4j} = \pi_{1j} + \pi_{2j} + \pi_{3j} + \pi_{4j}; \gamma_{5j} = \pi_{1j} + \pi_{2j} + \pi_{3j} + \pi_{4j} + \pi_{5j}; \gamma_{6j} = 1$$

$$\text{logit}(\gamma_{1j}) = \beta_0 \text{cons.}(\leftarrow ges_1)_{ij}$$

$$\text{logit}(\gamma_{2j}) = \beta_1 \text{cons.}(\leftarrow ges_2)_{ij}$$

$$\text{logit}(\gamma_{3j}) = \beta_2 \text{cons.}(\leftarrow ges_3)_{ij}$$

$$\text{logit}(\gamma_{4j}) = \beta_3 \text{cons.}(\leftarrow ges_4)_{ij}$$

$$\text{logit}(\gamma_{5j}) = \beta_4 \text{cons.}(\leftarrow ges_5)_{ij}$$

$$\text{cov}(\gamma_{sj}, \gamma_{rj}) = \gamma_{sj}(1 - \gamma_{rj}) / cons_j \quad s < r$$

Modell 39: Ordered Multinomial Modell

$$resp_{ij} \sim \text{Ordered Multinomial}(cons_j, \pi_{ij})$$

$$\gamma_{1j} = \pi_{1j}; \gamma_{2j} = \pi_{1j} + \pi_{2j}; \gamma_{3j} = \pi_{1j} + \pi_{2j} + \pi_{3j};$$

$$\gamma_{4j} = \pi_{1j} + \pi_{2j} + \pi_{3j} + \pi_{4j}; \gamma_{5j} = \pi_{1j} + \pi_{2j} + \pi_{3j} + \pi_{4j} + \pi_{5j}; \gamma_{6j} = 1$$

$$\text{logit}(\gamma_{1j}) = -3,587(0,050) \text{cons.}(\leftarrow ges_1)_{ij}$$

$$\text{logit}(\gamma_{2j}) = -2,621(0,033) \text{cons.}(\leftarrow ges_2)_{ij}$$

$$\text{logit}(\gamma_{3j}) = -2,077(0,026) \text{cons.}(\leftarrow ges_3)_{ij}$$

$$\text{logit}(\gamma_{4j}) = -1,241(0,020) \text{cons.}(\leftarrow ges_4)_{ij}$$

$$\text{logit}(\gamma_{5j}) = 0,278(0,017) \text{cons.}(\leftarrow ges_5)_{ij}$$

$$\text{cov}(\gamma_{sj}, \gamma_{rj}) = \gamma_{sj}(1 - \gamma_{rj}) / cons_j \quad s < r$$

Modell 40: Ordered Multinomial Modell - Schätzwerte

Daraus werden nun die *antilogits* in Tabelle 50 berechnet. Die Wahrscheinlichkeit, dass eine Person mit ‚nicht zufrieden‘ antwortet ist 2,69%. Die Wahrscheinlichkeit, dass eine Person mit ‚nicht zufrieden (1)‘ oder (2) antwortet ist 6,78%. Dieser Wert stimmt mit den %-werten der Vorkommnisse der einzelnen Kategorien in der Gesamtzufriedenheitsvariablen aus Tabelle 51 überein.

	Logit	Probability
N =	5	5
1	-3,5870	0,026936
2	-2,6210	0,067799
3	-2,0770	0,11135
4	-1,2410	0,22426
5	0,27800	0,56906

Tabelle 50: Anti-Logits

	1	2	3	4	5	6	TOTALS
N	399	606	646	1674	5109	6387	14821
%	2,7	4,1	4,4	11,3	34,5	43,1	100,0

Tabelle 51: Kategorienprozentwerte für Gesamtzufriedenheit

Nun wird ein Modell 41 mit Level 2 erstellt, welches aufgrund der Hierarchien mit den Kategorien zum Level-3-Modell wird. Eine Möglichkeit, Variation auf Kinostandortebene zuzulassen, ist jedem Kategorien-*intercept* Variation zu erlauben, was zu einer 5x5 Kovarianzmatrix auf Level 3 Ebene führt. Dies entspricht einem *simple multinomial two-level*-Modell.

$$resp_{ijk} \sim \text{Ordered Multinomial}(cons_{jk}, \pi_{ijk})$$

$$\gamma_{1jk} = \pi_{ij}, \gamma_{2jk} = \pi_{1jk} + \pi_{2jk}, \gamma_{3jk} = \pi_{1jk} + \pi_{2jk} + \pi_{3jk};$$

$$\gamma_{4jk} = \pi_{1jk} + \pi_{2jk} + \pi_{3jk} + \pi_{4jk}; \gamma_{5jk} = \pi_{1jk} + \pi_{2jk} + \pi_{3jk} + \pi_{4jk} + \pi_{5jk}; \gamma_{6jk} = 1$$

$$\log it(\gamma_{1jk}) = \beta_{0k} cons_{jk} (\Leftarrow ges_1)_{jk}$$

$$\beta_{0k} = \beta_0 + v_{0k}$$

$$\log it(\gamma_{2jk}) = \beta_{1k} cons_{jk} (\Leftarrow ges_2)_{jk}$$

$$\beta_{1k} = \beta_1 + v_{1k}$$

$$\log it(\gamma_{3jk}) = \beta_{2k} cons_{jk} (\Leftarrow ges_3)_{jk}$$

$$\beta_{2k} = \beta_2 + v_{2k}$$

$$\log it(\gamma_{4jk}) = \beta_{3k} cons_{jk} (\Leftarrow ges_4)_{jk}$$

$$\beta_{3k} = \beta_3 + v_{3k}$$

$$\log it(\gamma_{5jk}) = \beta_{4k} cons_{jk} (\Leftarrow ges_5)_{jk}$$

$$\beta_{4k} = \beta_4 + v_{4k}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \\ v_{3k} \\ v_{4k} \end{bmatrix} \sim N(0, \Omega_v); \Omega_v = \begin{bmatrix} \sigma_{v0}^2 & & & & \\ \sigma_{v01} & \sigma_{v1}^2 & & & \\ \sigma_{v02} & \sigma_{v12} & \sigma_{v2}^2 & & \\ \sigma_{v03} & \sigma_{v13} & \sigma_{v23} & \sigma_{v3}^2 & \\ \sigma_{v04} & \sigma_{v14} & \sigma_{v24} & \sigma_{v34} & \sigma_{v4}^2 \end{bmatrix}$$

$$cov(\gamma_{sjk}, \gamma_{rjk}) = \gamma_{sjk} (1 - \gamma_{rjk}) / cons_{jk} \quad s < r$$

Modell 41: Ordered Multinomial Modell auf Level-3-Ebene

Im Modell 42 wird ein *single variance* Term auf Kinostandortebene eingeführt, mit einheitlichen Koeffizienten für alle fünf Kategorien. Der zusätzliche Koeffizient h_{jk} ist als einfacher allgemeiner Koeffizient für alle Kategorien gleich. Es gibt jedoch auch die Möglichkeit, beliebige einzelne Kategorien mit ihren Koeffizienten zusammenzuhängen.

Es darf in einem Modell nicht gleichzeitig einen einheitlichen Koeffizienten für jede Kategorie und einen allgemeinen Koeffizienten geben. Nun wird die

allgemeine zwischenstandortbezogene Variabilität berücksichtigt und die Variable aus dem fixen Part entfernt. Die zweite und dritte Zeile spezifizieren das kumulative kategorische Modell. Für jede der kumulativen Kategorien folgen die Gleichungen für die fünf Antwortvariablen. Die erste Variable ist jeweils eine Konstante, welche dem *intercept* erlaubt, jeweils unterschiedliche Werte anzunehmen. Die weitere Variable ist eine Konstante mit durchgehendem Wert eins, welche dazu dient, durch den Zufallskoeffizienten, den gleichen Zufallsfehler jeder der fünf Kategoriengleichungen zuzuordnen. Somit wird eine einheitliche Kinostandortvarianz für jede Kategorie geschätzt.

$$\begin{aligned}
 resp_{ijk} &\sim \text{Ordered Multinomial}(cons_{jk}, \pi_{ijk}) \\
 \gamma_{1,jk} &= \pi_{ij}; \gamma_{2,jk} = \pi_{1,jk} + \pi_{2,jk}; \gamma_{3,jk} = \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk}; \\
 \gamma_{4,jk} &= \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk} + \pi_{4,jk}; \gamma_{5,jk} = \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk} + \pi_{4,jk} + \pi_{5,jk}; \gamma_{6,jk} = 1 \\
 \log it(\gamma_{1,jk}) &= \beta_0 cons.(\leq ges_1)_{jk} + h_{jk} \\
 \log it(\gamma_{2,jk}) &= \beta_1 cons.(\leq ges_2)_{jk} + h_{jk} \\
 \log it(\gamma_{3,jk}) &= \beta_2 cons.(\leq ges_3)_{jk} + h_{jk} \\
 \log it(\gamma_{4,jk}) &= \beta_3 cons.(\leq ges_4)_{jk} + h_{jk} \\
 \log it(\gamma_{5,jk}) &= \beta_4 cons.(\leq ges_5)_{jk} + h_{jk} \\
 h_{jk} &= \beta_6 cons.12345 \\
 cov(\gamma_{sjk}, \gamma_{rjk}) &= \gamma_{sjk}(1 - \gamma_{rjk}) / cons_{jk} \quad s \leq r
 \end{aligned}$$

Modell 42: Einführung eines single variance Terms

Modell 43 ist nicht mehr schätzbar.

$$\begin{aligned}
 resp_{ijk} &\sim \text{Ordered Multinomial}(cons_{jk}, \pi_{ijk}) \\
 \gamma_{1,jk} &= \pi_{ij}; \gamma_{2,jk} = \pi_{1,jk} + \pi_{2,jk}; \gamma_{3,jk} = \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk}; \\
 \gamma_{4,jk} &= \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk} + \pi_{4,jk}; \gamma_{5,jk} = \pi_{1,jk} + \pi_{2,jk} + \pi_{3,jk} + \pi_{4,jk} + \pi_{5,jk}; \gamma_{6,jk} = 1 \\
 \log it(\gamma_{1,jk}) &= \beta_0 cons.(\leq ges_1) + h_{jk} \\
 \log it(\gamma_{2,jk}) &= \beta_1 cons.(\leq ges_2) + h_{jk} \\
 \log it(\gamma_{3,jk}) &= \beta_2 cons.(\leq ges_3) + h_{jk} \\
 \log it(\gamma_{4,jk}) &= \beta_3 cons.(\leq ges_4) + h_{jk} \\
 \log it(\gamma_{5,jk}) &= \beta_4 cons.(\leq ges_5) + h_{jk} \\
 h_{jk} &= v_{6k} cons.12345 \\
 [v_{6k}] &\sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma^2_{v_6} \end{bmatrix} \\
 cov(\gamma_{sjk}, \gamma_{rjk}) &= \gamma_{sjk}(1 - \gamma_{rjk}) / cons_{jk} \quad s \leq r
 \end{aligned}$$

Modell 43: Grenzen der ordinalen Modellmöglichkeiten

Die *first order*-Schätzung würde die Parameter, verglichen mit der *second order*-Schätzung, unterschätzen. Erweiterungen und Abwandlungen daraus sind ebenfalls möglich. Es kann z.B. die Technik/Bild/Ton-Zufriedenheit mit

einheitlichem Koeffizienten als erklärende Variable eingeführt werden. Erlaubt man separate Koeffizienten für jede Kategorie, kommt dies einem multinomialen Modell gleich. Zusätzlich können weitere erklärende Variable mit quadratischem Term eingefügt werden.

4.3.4.12. Die Zeit als eigene Ebene - Kinodaten

In zeitübergreifenden Mehrebenenmodellen ist der Beobachtungszeitpunkt Level 1 und die Person Level 2. Der Kinostandort kann dabei als Level 3 fungieren. Es ist nicht notwendig, zu jeder Level 2 Person die vollständigen Level 1 Messungen im Datensatz zu haben. Trotzdem kann der gesamte vorhandene Datensatz verwendet werden. Die durchschnittliche Gesamtzufriedenheit des ab hier verwendeten Datensatzes über die zwölf Monate ist in Tabelle 52 dargestellt. Es zeigt sich eine sinkende Varianz in den späteren Beobachtungsmo- naten und ein etwa gleich bleibender Mittelwert.

	1	2	3	4	5	6	
N	3428	2989	3300	3009	2814	2348	
MEANS	4,95	4,98	4,98	5,02	4,88	4,95	
SD'S	1,33	1,30	1,27	1,24	1,30	1,24	
N	7	8	9	10	11	12	TOTALS
MEANS	2199	2461	2130	2968	2198	2027	31871
SD'S	4,93	4,87	4,98	4,98	4,98	5,07	4,96

Tabelle 52: Varianzveränderung

Modell 44 läuft über zwölf Zeitpunkte. Level 1 ist die jeweilige Beobachtung, Level 2 die jeweilige Person. Die erklärende Variable bei den wiederholten Messungen ist Zeitpunkt eins bis zwölf.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons$$

$$\beta_{0ij} = 5,3883240000(0,0073959390) + u_{0ij} + e_{0ij}$$

$$[u_{0ij}] \sim N(0, \Omega_u) : \Omega_u = [0,16890680000(0,0055826820)]$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [0,61424690000(0,0053274520)]$$

$$-2 * \log \text{likelihood}(IGLS \text{ Deviance}) = 79284,2800000000(31650 \text{ of } 67968 \text{ cases in use})$$

Modell 44: Einfaches longitudinales Modell

Die Variation zwischen den Zeitpunkten innerhalb der Probanden über die Zeit ist größer, im Vergleich zur Variation zwischen den Probanden.

4.3.4.13. Linear growth curve model - Kinodaten

Um die Variation zwischen den Beobachtungen zu erklären, die zwischen den Zeitpunkten und innerhalb der einzelnen Probanden liegt, die Level 1 Variation, wird ein fixer linearer Trend in einem *linear growth curve*-Modell 45 geschätzt. Dafür wird die Zeitvariable als fixe erklärende Variable in das Modell aufgenommen.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + -0,00811110230(0,0013154610)zeit_{ij} \\
 \beta_{0ij} &= 5,4362370000(0,0106945100) + u_{0j} + e_{0ij} \\
 \begin{bmatrix} u_{0j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = [0,1686288000(0,0055823910)] \\
 \begin{bmatrix} e_{0ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = [0,6135194000(0,0053271530)] \\
 -2 * \log likelihood(IGLS Deviance) &= 79246,5600000000(31650 \text{ of } 67968 \text{ cases in use})
 \end{aligned}$$

Modell 45: Linear Growth Curve Model

Der *likelihood*-Wert sinkt um $79.284,280 - 79.246,560 = 37,720$ bei einem zusätzlichen Parameter und ist hoch signifikant bei $8,1663e^{-10}$. Die Varianzen ändern sich kaum. Der durchschnittliche Zufriedenheitswert ist jedoch gestiegen, da ein Teil davon nun durch die negativen Zeitirritationen erklärt wird. In Modell 46 wird angenommen, dass die lineare Wachstumsrate von Proband zu Proband variiert und nicht fix ist. Dies wurde nämlich im vorangegangenen Modell angenommen. Der Koeffizient der Zeitvariablen ist zufällig auf Level 2. Die Devianz beträgt $79.246,560 - 79.185,550 = 61,01$, verglichen mit dem vorangegangenen Modell und bei zwei zusätzlichen Parametern hoch signifikant mit $0,00000$. Somit können zu berücksichtigende Unterschiede der linearen Wachstumsraten zwischen den Probanden festgestellt werden.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + \beta_{1j}zeit_{ij} \\
 \beta_{0ij} &= 5,4369340000(0,0104924700) + u_{0j} + e_{0ij} \\
 \beta_{1j} &= -0,0082739240(0,0013893670) + u_{1j} \\
 \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,1495974000(0,0110724300) \\ -0,0009079265(0,0012451300) \quad 0,0008431631(0,0001831000) \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = [0,6033836000(0,0056300640)] \\
 -2 * \log likelihood(IGLS Deviance) &= 79185,5500000000(31650 \text{ of } 67968 \text{ cases in use})
 \end{aligned}$$

Modell 46: Probandenvariation der Wachstumsveränderung

Die Quadratwurzel von σ^2_{u1} ist $\sqrt{0,0008431631} = 0,02903727$ und ergibt die Standardabweichung. Geht man von normalverteilten Daten aus, so haben 96 %

der Probanden Wachstumsraten innerhalb der doppelten Standardabweichung über und unter dem Gesamtmittelwert von $-0,008273924$. Dies ergibt Intervalle von $-0,008273924+2*0,02903727=0,049800616$ und $-0,008273924-2*0,02903727=-0,066348464$. Diagramm 108 zeigt die standardisierten Level-2-Residuen.

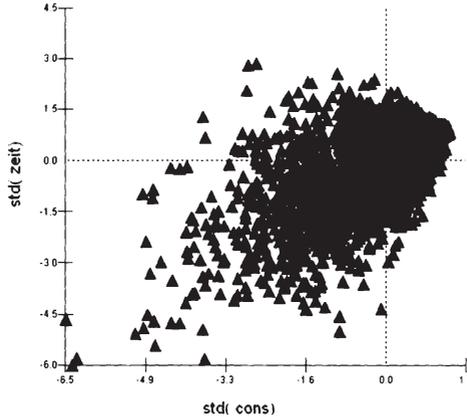


Diagramm 108: Standardisierte Level-2-Residuen

Die errechnete Korrelation der beiden Level 2 Residuen liegt lediglich bei

$\frac{-0,0009079265}{\sqrt{0,1495974*0,0008431631}} = -0,0816$. Die niedrige Korrelation zwischen den Zufriedenheitslevelresiduen und den zeitlichen Veränderungsresiduen bedeutet, dass zwischen den Zufriedenheitslevels und den Zufriedenheitsveränderungen kein Zusammenhang besteht.

Die Level-1-Residuen gegen die Normalverteilungswerte in Diagramm 109 geplottet sind ebenfalls nicht perfekt verteilt.

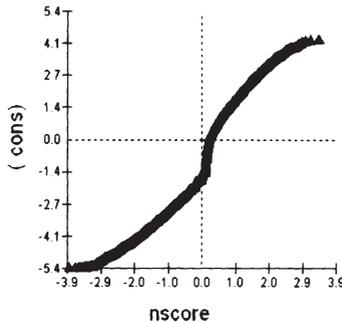


Diagramm 109: Residuenverteilung

Im Modell 47 wird eine nicht-konstante Level-1-Variation eingefügt, welche die konstante Koeffizientenvariation der Gesamtzufriedenheit darstellt. Somit wird über die Zeit hinweg von personenspezifischen Entwicklungen und darin innewohnenden Varianzen ausgegangen. Dafür muss die Gesamtvarianz der Technik/Bild/Ton-Zufriedenheiten proportional zum Quadrat des Mittelwertes sein, um nicht konstante Level-1- und Level-2-Variation erwarten zu können. Um die Level-1-Varianz als quadratische Funktion des vorhergesagten Wertes darstellen zu können, muss der Technik/Bild/Ton-Koeffizient zufällig auf Level 1 sein.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + \beta_{1ij}zeit_{ij} \\
 \beta_{0ij} &= 5,4369300000(0,01049245800) + u_{0j} + e_{0ij} \\
 \beta_{1ij} &= -0,0082734210(0,0013893890) + u_{1j} + e_{1ij} \\
 \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,1496128000(0,0110729900) & \\ -0,0009101765(0,0012454210) & 0,00084435328(0,0001831711) \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0,6033757000(0,0056300740) & \\ 0,0000000000(0,0000000000) & 0,0000000000(0,0000000000) \end{bmatrix} \\
 -2 * \log \text{likelihood} (\text{IGLS Deviance}) &= 79185,59000000000(31650 \text{ of } 67968 \text{ cases in use})
 \end{aligned}$$

Modell 47: Level-1-Variation

Die Varianzfunktion auf Level 1 ist $var(e_{0ij}x_0 + e_{1ij}x_{1ij}) = \delta_{e0}^2 x_0^2 + 2\delta_{e01}x_0x_{1ij} + \delta_{e1}^2 x_{1ij}^2$. Auf Level 1 gibt es keine Variation über die Beobachtungspunkte, da die Messpunkte mit der Zeitvariablen übereinstimmen.

4.3.4.14. Nicht-Linearitäten - Kinodaten

Das Wachstum der Gesamtzufriedenheit ist möglicherweise nicht linear über die Zeit hinweg. Um nicht-lineares Wachstum einzubringen, wird in den Modellen 48, 49 und 50 als zusätzliche Variable ein quadratischer Term, nämlich das Quadrat der Variablen Zeit berechnet und eingebracht.

$$\begin{aligned}
 ges_{ij} &\sim N(XB, \Omega) \\
 ges_{ij} &= \beta_{0ij}cons + \beta_{1ij}zeit_{ij} + 0,001(0,000)zeitsq_{ij} \\
 \beta_{0ij} &= 5,470(0016) + u_{0j} + e_{0ij} \\
 \beta_{1ij} &= -0,024(0,006) + u_{1j} \\
 \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,149(0,011) & \\ -0,001(0,001) & 0,001(0,000) \end{bmatrix} \\
 \begin{bmatrix} e_{0ij} \end{bmatrix} &\sim N(0, \Omega_e) : \Omega_e = [0,603(0,006)] \\
 -2 * \log \text{likelihood} (\text{IGLS Deviance}) &= 79177,750(31650 \text{ of } 67968 \text{ cases in use})
 \end{aligned}$$

Modell 48: Nicht-lineares Modell ohne quadratischer Level-2-Variation

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons + \beta_{1ij} zeit_{ij} + \beta_{2ij} zeitsq_{ij}$$

$$\beta_{0ij} = 5,470(0,016) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = -0,024(0,006) + u_{1j}$$

$$\beta_{2ij} = 0,001(0,000) + u_{2j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,132(0,024) & & \\ -0,005(0,008) & 0,009(0,003) & \\ 0,000(0,001) & -0,001(0,000) & 0,000(0,000) \end{bmatrix}$$

$$[e_{0ij}] \sim N(0, \Omega_e) : \Omega_e = [0,598(0,006)]$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 79140,480 (31650 \text{ of } 67968 \text{ cases in use})$$

Modell 49: Nicht-lineares Model mit quadratischer Level-2-Variation

Der likelihood-Wert sinkt um $79.185,590 - 79.140,480 = 45,110$ bei 4 Freiheitsgraden, ein fixer und vier zufällige Parameter, und ist hoch signifikant mit $3,7901e^{-9}$. Somit variiert der quadratische Term zwischen den Probanden und verbessert durch Einbringung der Zeitvariablen die Genauigkeit der Modellschätzungen. Von Modell zu Modell ist das Level-1-Residuum gesunken.

$$ges_{ij} \sim N(XB, \Omega)$$

$$ges_{ij} = \beta_{0ij} cons + \beta_{1ij} zeit_{ij} + \beta_{2ij} zeitsq_{ij}$$

$$\beta_{0ij} = 5,470428(0,015759) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = -0,023649(0,005797) + u_{1j} + e_{1ij}$$

$$\beta_{2ij} = 0,001221(0,000444) + u_{2j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0,131795(0,024062) & & \\ -0,005098(0,008119) & 0,008641(0,003227) & \\ 0,000169(0,000592) & -0,000556(0,000241) & 0,000040(0,000019) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim N(0, \Omega_e) : \Omega_e = \begin{bmatrix} 0,597597(0,005981) \\ 0,000000(0,000000) & 0,000000(0,000000) \end{bmatrix}$$

$$-2 * \text{loglikelihood (IGLS Deviance)} = 79140,460000 (65016 \text{ of } 67968 \text{ cases in use})$$

Modell 50: Nicht-lineares Modell mit Level-1-Variation

Nun werden die Level-1 und 2-Varianzen summiert. Dies ergibt die gesamte vorhergesagte Varianz und diese wird über die Zeit in Diagramm 110 abgeplottet. Es ist erkennbar, dass die Varianz bis zur achten Beobachtung ansteigt, aber bis Monat zwölf wieder abfällt. Dies könnte dadurch begründet werden, dass zuerst alle Probanden zufrieden sind, in den Folgemonaten jedoch aufgrund der ständigen Konfrontation mit dem Kinothema und der Beschäftigung damit kritischer und ausdifferenzierter werden. Später entwickeln sie

wieder ein klareres Antwortmuster und das Modell erklärt die Varianz wieder besser.

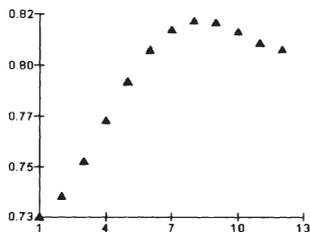


Diagramm 110: Zeitliche Varianzentwicklung

Die individuellen Probandenveränderungen sehen wie in Diagramm 111 aus. Die Kurven weisen verschiedene Trends über die zwölf Monate hinweg auf. Jeweils nur für die Beobachtungszeitpunkte werden die Werte geschätzt. Somit gibt es schöne gekrümmte Kurven auch nur für jene Personen, welche in etwa durchgehend zu jedem Beobachtungszeitpunkt ein Kino bewertet haben und Daten zur Berechnung vorhanden sind. Es sind Probanden erkennbar, für welche Bewertungseinstieg und Abbruch über die Befragungswellen deutlich erkennbar sind. Auch unterschiedlich gekrümmte Kurven bzw. Steigungen der Kurvenverläufe treten auf.

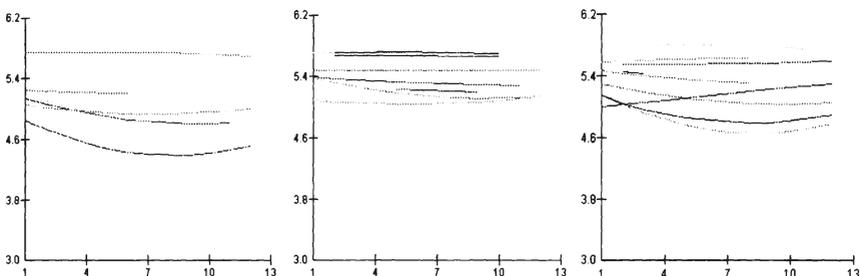


Diagramm 111: Individuelle Personenentwicklungen

Diagramm 112 plottet alle Probanden zusammen. Es finden sich am unteren Ende der Graphik einige wenige Ausreißer. Insgesamt werden Informationen zu 31.659 Bewertungen über zwölf Monate von 5.664 in Summe befragten Personen verwendet. Ein Proband hat durchschnittlich 5,5895 Besuche nach deren Gesamtzufriedenheit bewertet. Der Großteil der Verläufe befindet sich in den besseren Zufriedenheitsbereichen in der Umgebung der Bewertungen ‚sehr zufrieden (6)‘ und (5).

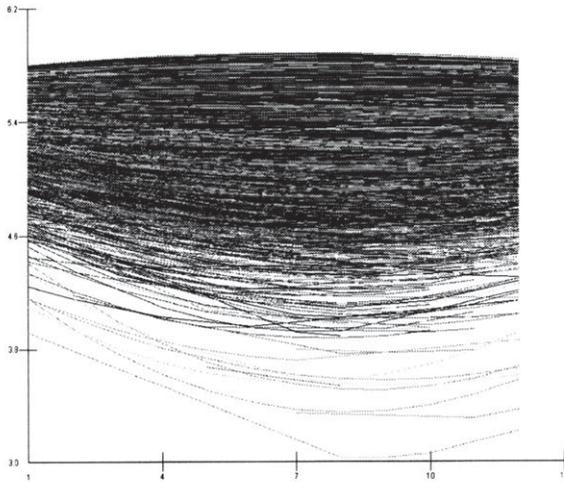


Diagramm 112: Zeitliche Veränderung alle Probanden

In Diagramm 113 zeigt die horizontale x-Achse in allen Graphiken den gleichbleibenden individuellen *intercept*, nämlich die unzufriedenen Technik/Bild/Ton-Zufriedenheitsbewertungen auf der linken Seite und die sehr zufriedenen auf der rechten. Dieser Level steht für den zeitinvarianten Zufriedenheitslevel jeder Person. Die y-Achse in der linken Graphik beinhaltet die Zeitpunkte, im Vordergrund Zeitpunkt eins und im Hintergrund den letzten Beobachtungszeitpunkt in Welle zwölf. In der mittleren und rechten Graphik wird dies durch die vertikale Achse dargestellt. Die vertikale z-Achse in der linken Graphik beschreibt die Anzahl an Bewertungen. In der mittleren und rechten Graphik wird die Beobachtungsanzahl durch die Farbdichte beschrieben.

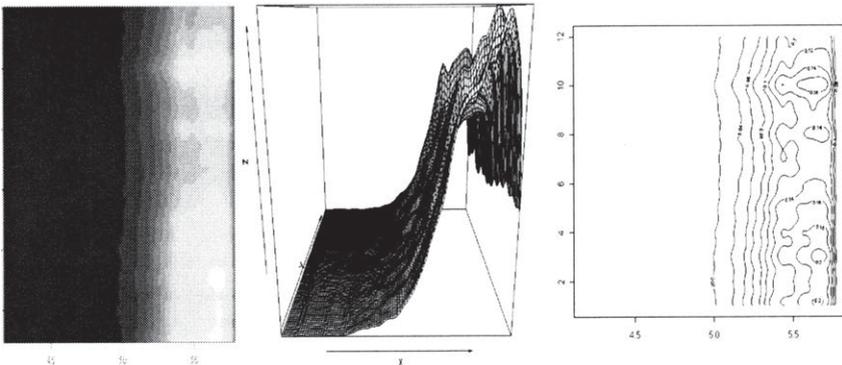


Diagramm 113: Contour Plots longitudinale Zufriedenheitsbeobachtungen

Es existieren viele Beobachtungen in den höheren Zufriedenheitslagen, speziell zu den Zeitpunkten eins, zwei, drei, acht und zehn. Weniger Beobachtungen gibt es bei Zufriedenheitswerten mit niedrigen intercepts.

Diagramm 114 gibt einen Überblick der individuellen zeitabhängigen Gesamtzufriedenheitstrends.

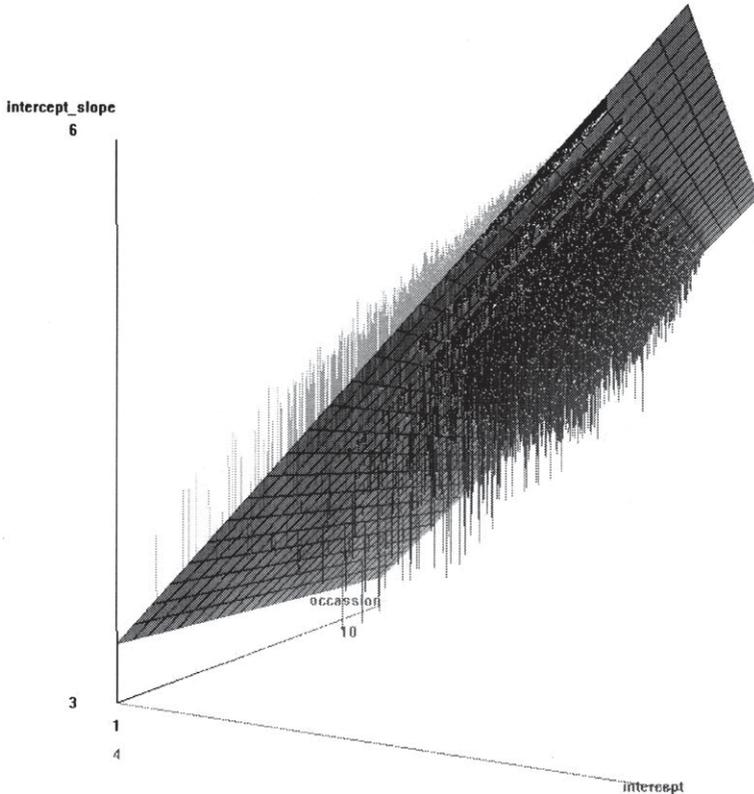


Diagramm 114: Zeitlich individuelle Varianzen

6

Die *intercept*-Achse als zeitinvarianter Gesamtzufriedenheitslevel und der Beobachtungszeitpunkt werden wie zuvor dargestellt. Jeder Proband befindet sich zu jedem Beobachtungszeitpunkt auf dem gleichen *intercept*-Level auf der *intercept*-Achse. Die vertikale Achse, *intercept_slope*, ist die Summe der individuellen *intercept*-Werte auf der *intercept*-Achse, plus die *slope*-Variation

zum jeweiligen Zeitpunkt, also die zeitabhängige Zufriedenheitsveränderung des Individuums. Nur die zeitabhängigen Aspekte werden dargestellt und der zeitunabhängige *intercept* dient lediglich als Startwert. Die Fläche wird durch Regression zwischen *intercept* und *intercept_slope* definiert. Sie stellt den zeitunabhängigen mittleren Zufriedenheitslevel aller Probanden und Beobachtungen dar. Die vertikalen Linien oberhalb der Fläche stellen positive und die darunter negative Abweichungen dar. Das Ende jeder Linie entspricht dem geschätzten individuellen Zufriedenheitswert zu den einzelnen Zeitpunkten. Sie repräsentieren also die zeitabhängigen Zufriedenheitslevels der Individuen.

Die meisten vorhergesagten Zufriedenheitswerte befinden sich auf den höheren Zufriedenheitslevels. Es gibt keine Werte unter drei auf der *intercept_slope*-Achse und keine Werte niedriger als vier auf der *intercept*-Achse. Folglich existieren sinkende Zufriedenheitstrends über die Zeit, da die niedrigsten zeitvarianten Werte niedriger als die niedrigsten zeitinvarianten Werte liegen. Dieser Effekt resultiert auch aus den modellbasierten angenommenen Verteilungen.

Für die höheren *intercept*-Werte, welche zum Zeitpunkt eins Richtung obere rechte Ecke der Fläche starten, zeigen sich steigende Werte. Die negativen Variationen unter der Regressionsfläche verschwinden über die Zeit. Gründe dafür liegen in den gesteigerten Zufriedenheitswerten, welche zu einem Level von sechs, *delight*, führen. Der Zufriedenheitslevel kann jedoch nicht gehalten werden, was auf einen generell sinkenden Zufriedenheitstrend aufgrund der Adaption der Qualitätslevel hindeutet.

Um unterschiedliche Zufriedenheitssegmente aufzudecken, werden die Individuen über *k-means* Clustering gruppiert und deren Verlauf und Variation in Diagramm 115 dargestellt. Die Lösung mit vier Clustern zeigte die am deutlichsten unterschiedlichen Cluster. Jede weitere Erhöhung der Clusteranzahl teilt lediglich den unzufriedenen Cluster in weitere Cluster auf, führt jedoch zu keinen wesentlichen Interpretationsunterschieden. In der linken Graphik werden die Zufriedenheitsverläufe der Cluster über die Zeit visualisiert. In der rechten Graphik finden sich die Probandenzahlen der vier Cluster und die gruppierte Visualisierung aus Diagramm 114.

Aus der rechten Graphik kommt deutlich zum Vorschein, wie auch schon in vorangehenden Analysen gezeigt, dass die Variation auf den unzufriedenen Levels, verglichen mit den zufriedenen Levels, höher ist.

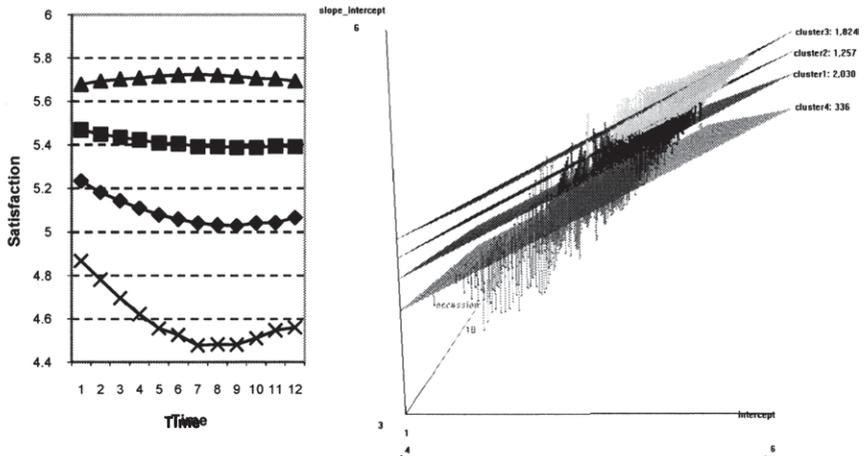


Diagramm 115: Clustering über die Zufriedenheitsverläufe

Weniger zufriedene Besucher weisen eine höhere Varianz auf, wie auch in Diagramm 116 gezeigt wird. Sie müssen somit auch gruppenweise unterschiedlich behandelt werden, um die Heterogenität zu berücksichtigen.

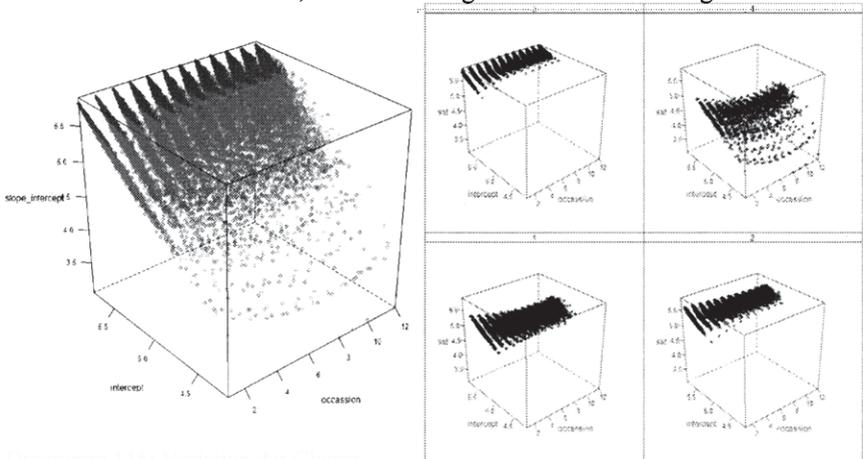


Diagramm 116: Variation der Cluster

4.3.4.15. Übergang von der regressionsbasierten zur kovarianzbasierten Lösung

Stoel et al. (2003) beschreiben die Unterschiede zwischen *multilevel regression* (MLR) und *latent growth curve* (LGC) Modellen. Beide Verfahren untersuchen intraindividuelle als auch interindividuelle Differenzen der Veränderungen. MLR

basiert auf hierarchischer linearer Regression und LGC auf SEM. In MLR wird die Zeit als unabhängige fixe Variable auf dem ersten Level gehandelt, also als zusätzliche Variable eingebracht und die Person auf dem zweiten Level, wobei auf jedem Level erklärende Variablen einwirken können und die inter-individuellen Unterschiede in den Parametern der Wachstumskurve als *random effects* verstanden werden. Der *initial level* und der *slope* werden durch Koeffizienten repräsentiert. In den LGC Modellen wird die Zeit als latente Variable und somit als Dimension gehandelt. Der *initial level* und *slope* wird durch latente Faktoren bestimmt. Die Zeit wird durch Faktorenladungen eingebracht. Die Parameter der individuellen Kurven sind latente Variablen mit einer Kovarianz und einer Mittelwertstruktur, gegengleich zu den *random effects* in MLR. Durch die unterschiedliche Einbringung der Zeit ist MLR ein univariates Verfahren, wobei Zeitpunkte als Beobachtungen der gleichen Variablen gehandelt werden. LGC Modelle sind multivariate Verfahren, wobei jeder Zeitpunkt als eigene Variable gehandelt wird.

Üblicherweise werden beide Analyseverfahren über ML geschätzt, welcher durch Maximierung der *likelihood*-Funktion iterativ Schätzungen für die Populationsparameter berechnet, die die Wahrscheinlichkeit der Beobachtung der Daten gegeben das Modell maximieren und asymptotisch effiziente und konsistente Schätzungen liefert. Andere Schätzverfahren sind *restricted-ML* (RML), Markov Chain Monte Carlo (MCMC) oder *bootstrapping* Methoden.

Ein Vorteil von LGC ist die Einbringung mehrerer Indikatoren in ein Modell höherer Ordnung. Die Variation der multiplen Indikatoren wird durch die Faktoren erster Ordnung aufgefangen und die Faktoren zweiter Ordnung erklären den Mittelwert und die Kovarianzstruktur der Faktoren erster Ordnung. In MLR müssten dafür Summen der Indikatoren oder die Berechnung der Wachstumsverläufe für die einzelnen Indikatoren verwendet werden, wobei die Verbindung einzeln gemessener Wachstumsverläufe ein Problem darstellt.

SEM ermöglicht auch den Test auf Invarianz des untersuchten latenten Konstruktes, welches durch gleiche Faktorenladungen über die Zeit inhaltlich gleich gehalten wird. Die genesteten Modelle mit vs. ohne Restriktion werden über den χ^2 -Differenztest verglichen. Zeigen sich signifikante Abweichungen, wird davon ausgegangen, dass die gemessenen Indikatoren über die Zeit nicht das gleiche Konstrukt messen. In MLR sind die Restriktionen vorgegeben und können nicht aufgeweicht bzw. getestet werden.

MLR als auch LGC können über polynomiale nicht-lineare Verläufe schätzen. In LGC Modellen wird durch freie Schätzung der Wachstumsfaktorkoeffizienten der Verlauf optimal geschätzt. In MLR ist das nicht möglich. Durch Fixierung der Zeitvariablen in MLR, welche zuvor über LGC optimiert wurden, zeigen die Modelle den gleichen LR-Test-Vergleich des linearen und nicht-linearen Modells.

LGC ermöglicht, durch die Behandlung der Beobachtung als eigene Variable, eine separate Schätzung von Varianzen und Kovarianzen. Auch der Einfluss von *time-varying* Kovariaten kann für jede Beobachtung unterschiedlich geschätzt werden. MLR erlaubt dies teilweise über eine komplizierte Einführung von Dummyvariablen. Beide Modelle liefern gleiche Parameterschätzungen.

In MLR ist es einfacher mehrere Levels einzuführen. Wird dies ignoriert, wird die Annahme unabhängig und gleich verteilter Beobachtungen verletzt, da Gruppen auf einem Level möglicherweise unterschiedliche Eigenheiten aufweisen. Die Standardfehler werden nach unten verzerrt und es können sich fälschlicherweise signifikante Effekte ergeben. In LGC Modellen sind derartige Analysen schwerer zu spezifizieren.

In MLR sind die Beobachtungszeitpunkte und der Abstand von Individuum zu Individuum gleichgültig und ermöglicht eine einfache Handhabung von fehlenden Werten. Fehlende Werte erklärender Variablen können nicht behandelt werden, nur die von den abhängigen. Dieses Problem besteht bei LGC nicht. LGC ermöglicht Lösungen über *full information maximum likelihood* (FIML), die alle Daten verwendet und durch Sortierung der Daten in Antwortpatterns und Verwendung eines *multiple group*-Designs und Gruppenbeschränkungen das gleiche Modell für alle Gruppen schätzt. FIML benötigt nicht die *missing completely at random* (MCAR), aber *missing at random* (MAR) Annahme. Letztere geht von der Vorhersagemöglichkeit fehlender Werte mit verfügbaren Daten aus.

SEM ermöglicht die einfache Schätzung von Mittelwerten und Kovarianzen, welche mit den latenten Wachstumsfaktoren verbunden sind, also direkte Beziehungen zwischen Wachstumsprozessen, simultane Beziehungen dieser Prozesse mit Kovariaten und Mediatoreffekte von Kovariaten.

Steele (2008) stellt Erweiterungen des *multilevel*-Ansatzes vor. Allgemein werden beobachtete Messungen in longitudinalen Analysen als zeitabhängig gesehen. In autoregressiven Modellen ist der *outcome* zum Zeitpunkt t eine

Funktion von verzögerten *outcomes*, wie beispielsweise der *outcome* zum Zeitpunkt $t-1$ in einem Modell erster Ordnung. Sie stellt Wachstumsmodelle vor, welche geeignet sind für autokorrelierte Residuen, Faktorenanalysemodelle für multiple Indikatoren, *event history*-Modelle, auch *survival* oder *duration analysis* genannt, Modelle für *competing risks* und *multiple states*, *multiple-change*, *event processes* und Wachstumskurvenmodelle, welche Autokorrelation beinhalten, mehrere Indikatoren in einem Messmodell und multivariate Antworten.

Es werden Spezialfälle der *multilevel*-Modelle beschrieben. Die Zeit wird dabei als Kovariate behandelt, wobei die Residuen als normalverteilt angenommen werden und Korrelation zwischen Residuen auf gleichem Level erlaubt sind. Positive *between-individual*-Varianz führt zu steigender also divergierender Varianz und negative zu konvergierender Varianz. Es kann passieren, dass sich die Wachstumsverläufe schneiden und die Varianz wiederum steigt.

Hierarchische Ebenen über der Individualebene werden in SEM über *multiple-group*-Analysen gelöst und für fehlende Werte stehen FIML und *multiple imputation* bereit. Die Vorzüge der Mehrebenenmodelle sind aufgrund der Generalisierbarkeit hierarchischer und nicht-hierarchischer Ansätze, nicht normalverteilter Antworten und gemischter Antworttypen in multivariaten Analysen jedoch nicht zu vernachlässigen. Sie werden durch eine kubische polynomiale Funktion dargestellt, wobei *random coefficients* für die vier Terme, intercept, linear, quadratisch und kubisch, erlaubt werden. Die Modelle werden über LR-Tests verglichen.

Da Beobachtungen von Individuen, welche näher beieinander liegen, ähnliche Abweichungen der individuellen Wachstumstrajektorie aufweisen und diese mit der Größe des Zeitintervalls steigen, bietet sich die Einführung von Autokorrelation an. Dies führt zu den *multilevel time series*-Modelle, welche Steele (2008) durch *discrete time first-order-auto-regressive* AR (1)-Modelle beschreibt.

Wiederholte latente Variable mit multiplen Indikatoren werden mit der latenten Variablen über ein Faktoren- oder Messmodell verbunden. Residuen auf Individuen und Beobachtungsindividuenlevel werden als unkorreliert über die Indikatoren angenommen. Durch das Strukturmodell gelangt man zum *multilevel* SEM. Dies stellt *multiple indicators growth curve*-Modelle dar, wobei die Veränderung als Stufenfunktion mit erklärenden Variablen (Dummyvariablen) eingeführt wird und individuelle statt linearer Veränderungen erlaubt sind.

Auch Kausalmodelle für multivariate Antworten, nämlich zwei *outcome*-Variablen, werden reziprok in einem *multilevel bivariate response*-Modell verbunden. Jeder der beiden Prozesse hängt vom verzögerten *outcome* des andern Prozesses ab. Eine *between-process* Residuenkorrelation erlaubt Korrelationen zwischen *time varying* oder individuenspezifischen nicht beobachtbaren Variablen, welche den Prozess beeinflussen und auch zwischen allen möglichen *random effects* auf gleichem Level. In dieser hierarchischen Struktur mit drei Ebenen sind die Antworten, Level 1, genestet innerhalb einer Messbeobachtung, Level 2, innerhalb des Individuums, Level 3. Das Modell kann auch als konfirmatorisches Faktorenmodell mit Faktoren als *random effects* angesehen werden. Im *cross-lagged*-Modell wird zusätzlich ein autoregressiver Term eingefügt. Werden sowohl das latente Wachstum als auch der autoregressive Term eingefügt, so führt dies zu den auto-regressiven Trajektorienmodellen. Auch Modelle, in welchen der *intercept* eines Wachstumsprozesses den *slope* eines anderen beeinflusst und eine höhere Gruppenebene dabei kontrolliert wird, werden in multilevel SEM vorgestellt, welche dann auch die Messfehlerkontrolle erlauben.

Weitere Modelle, welche das Auftreten eines Events messen werden auch aufgeführt. Theoretisch kann das Event zu jedem Zeitpunkt passieren. Deshalb wird die Zeit als diskret behandelt. Ein weiterer Grund die vorgestellten Modelle als *discrete time*-Modelle zu spezifizieren, liegt in der leichteren Erweiterung für wiederholte Events, *competing risks*, *multiple states* und *multiple processes*. Im *discrete time event history*-Modell wird ein einzelner Event beschrieben, im *multilevel event history model for repeated events* gibt es mehrere Events. Die neubeginnenden Episoden nach Auftreten eines Events, Level 1, sind innerhalb der Individuen, Level 2, genestet und werden mit *random effects logit*-Modellen oder *shared frailty*-Modellen gelöst. *Competing risks* bedeutet, dass eine Episode in unterschiedlichen Entwicklungen oder Arten von Events enden kann. Diese Modelle werden über *multinomiale logit*-Modelle gelöst. Auch Entwicklungen zwischen *multiple states* sind simultan modellierbar. In den *causal event history*-Modellen, auch *multi process*-Modell genannt, tauchen bivariate wiederholte Messungen auf, wobei Entwicklungen eines vorangehenden Prozesses die Terminierung eines anderen Prozesses bestimmen.

Zusammenfassend weisen Steele (2008) und Stoel et al. (2003) auf das Ineinanderfließen des *multilevel*- und SEM-Ansatzes hin. Auf *multilevel*-Seite betrifft dies die *multilevel factor*-Analyse, wo um *mixtures* und strukturelle Abhängigkeiten erweitert wird, auf SEM-Seite die Einführung hierarchischer Strukturen über *multiple-group*-Analysen. Beide Modelltypen können im

generalized linear latent and mixed modelling framework eingebettet werden und wurden in *gllamm* in Stata implementiert. Die Modelle unterscheiden nicht zwischen *random effects* in *multilevel*-Modellen und Faktoren in SEM, aber ermöglichen Flexibilität in der Spezifizierung der Ladungen der latenten Variablen. Ein *multilevel random effect* wird durch Definition einer latenten Variablen mit allen Ladungen auf eins fixiert und ein allgemeiner Faktor angepasst, wobei zumindest eine Ladung frei geschätzt wird.

4.3.5. Faktorenermittlung - Kinodaten

Ziele:

Die inhaltlichen Zusammenhänge der elf Zufriedenheitsvariablen werden anhand einer explorativen Faktorenanalyse analysiert.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den Monaten Oktober bis März);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Programme:

SPSS; Mplus;

Modell:

Sämtliche Zufriedenheitsbewertungen der ersten sechs Wellen werden zusammengeführt und Veränderungen über die Zeit außer Acht gelassen. Daraus resultieren 18.309 Bewertungen. Die Faktoren werden über die *principal component analysis* (PCA), Hauptkomponentenanalyse, ermittelt.

Zwecks leichterer Interpretation der Ergebnisse wird in der SPSS-Lösung die Rotationsmethode Varimax verwendet. Die sinnvollste Lösung zeigt eine Unterteilung in drei Faktoren aus Tabelle 53. Die drei Faktoren erklären zusammen 58,531 % der Varianz. Der erste Faktor beeinflusst verstärkt die Items Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten und Erscheinungsbild/Image und beschreibt 22,914 % der Varianz. Der dritte Faktor, mit 14,513 %-iger Varianzerklärung beeinflusst maßgeblich die Items Technik/Bild/Ton und Komfort. Der zweite Faktor mit einer Varianzerklärung von 21,105 %, beeinflusst hauptsächlich die Items

Servicepersonal, Preis/Leistung, Buffet und Atmosphäre. Die drei Dimensionen widerspiegeln den zeitlichen Ablauf eines Kinobesuches. Faktor eins steht für ‚vor dem Kinobesuch‘, Faktor zwei für ‚im Kino‘ und Faktor drei für ‚im Kinosaal‘. Dieser Faktorenaufbau wird für die noch folgenden Modelle verwendet.

Faktoren	SPSS – Varimax rotiert					
	1	2	3			
Technik/Bild/Ton	0,155	0,099	0,845			
Komfort	0,185	0,316	0,723			
Servicepersonal	0,209	0,659	0,316			
Preis/Leistung	0,053	0,708	0,206			
Buffet	0,191	0,790	-0,004			
Atmosphäre	0,492	0,602	0,121			
Standort/Erreichbarkeit	0,655	0,159	-0,039			
Filmangebot	0,718	-0,002	0,285			
Informationsangebot	0,702	0,189	0,280			
Reservierungsmöglichkeiten	0,575	0,163	0,067			
Erscheinungsbild/Image	0,607	0,450	0,190			
Faktoren	MPlus – ordinalskaliert			MPlus - intervallskaliert		
	1	2	3	1	2	3
Technik/Bild/Ton	0,286	0,181	0,669	0,140	0,212	0,590
Komfort	0,238	0,334	0,675	0,299	0,217	0,605
Servicepersonal	0,252	0,599	0,364	0,555	0,225	0,326
Preis/Leistung	0,154	0,543	0,251	0,501	0,133	0,239
Buffet	0,219	0,697	0,113	0,649	0,191	0,100
Atmosphäre	0,445	0,639	0,185	0,594	0,446	0,141
Standort/Erreichbarkeit	0,509	0,263	0,115	0,221	0,440	0,075
Filmangebot	0,714	0,113	0,252	0,075	0,637	0,228
Informationsangebot	0,725	0,255	0,255	0,212	0,663	0,249
Reservierungsmöglichkeiten	0,507	0,255	0,169	0,209	0,400	0,131
Erscheinungsbild/Image	0,563	0,507	0,240	0,470	0,548	0,187

Tabelle 53: Faktorenladungen

Unter Verwendung von Mplus zeigt sich ein ähnliches Bild für die ordinalskalierte, als auch für die intervallskalierte Lösung (10.626 Beobachtungen). Lediglich das Item Erscheinungsbild/Image wird verstärkt sowohl durch die Dimension ‚vor dem Kinobesuch‘ als auch ‚im Kino‘ geprägt. Dies kann möglicherweise dadurch erklärt werden, dass das Imagekonstrukt stärker in die Entscheidung vor dem jeweiligen Film fällt, das Erscheinungsbild jedoch verstärkt während des Kinobesuches zum Tragen kommt.

4.3.6. *Growth Curve Modelling: vom latent growth curve model zum finite growth mixture model*

Ziele:

Untersucht wird die Veränderung von Zufriedenheitsdimensionen über die Zeit und deren personenspezifische Segmentierung.

Variablen:

Wie beurteilen Sie Ihre Zufriedenheit mit diesem Kino in Bezug auf ...? (gemessen in den Monaten Oktober 2006 bis September 2007);

Items (#11): Technik/Bild/Ton, Komfort, Servicepersonal, Preis/Leistung, Buffet, Atmosphäre, Standort/Erreichbarkeit, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten, Erscheinungsbild/Image;

Skalierung: 6-stufige-Likert-Skala: 6 (sehr zufrieden) bis 1 (nicht zufrieden), -1 (keine Erfahrung);

Programm:

M-Plus;

Modell:

Preacher et al. (2008) sehen die Besonderheiten der latenten Wachstumsmodelle in der Verbindung sowohl struktureller als auch dynamischer Analysen im Sinne von Wachstums- und Entwicklungsmustern. LGC Modelle sind eine Mischung aus *variable-centered (nomothetic)* Modellen, wobei die Mittelwerte über die Zeit hinweg untersucht werden, und *person-centered (idiographic)* Modellen, wobei Veränderungen von Individuen untersucht werden. So werden Veränderungen latenter Konstrukte und nicht nur direkt beobachteter Indikatoren, welche eine Untergruppe dieser Analyseverfahren darstellen, untersucht. Die Modelle werden auch *multiple indicator growth model*, *curve of factors model*, *second order latent growth model*, *latent variable longitudinal curve model* oder *multivariate latent growth model* genannt. In den Modellen werden Daten mehrerer Items, welche über mehrere Zeitpunkte hinweg gemessen wurden, zu latenten Konstrukten zusammengefasst. Dann wird untersucht, ob diese latenten Konstrukte Veränderungen unterzogen sind. Aber auch die Entwicklung einzelner Items kann untersucht werden. Die Gewichte, mit welchen die Konstrukte auf den beobachteten Items laden, können über die Zeit hinweg als variabel angesehen werden. So besteht die Möglichkeit einer partiellen bzw. konfigurativen Faktorinvarianz. Auch autoregressive Stabilitäts- bzw. *simplex*-Modelle, in welchen latente Prozesse der Einstellungsveränderung gemessen werden, sind denkbar und werden später in dieser Arbeit auch

durchgeführt. Für die Veränderungsmessung werden zumindest zwei latente Wachstumsfaktoren erschaffen. Der Wachstumsfaktor *level* gibt das Basisniveau an, welches über den gesamten Zeitraum gleich bleibt. Der Wachstumsfaktor *slope* gibt die positive oder negative Abweichung an, welche im linearen Modell konstant, oder im kurvilinearen Modell, ohne zuvor fixiert zu werden, über die Zeit geschätzt wird. Intraindividuelle Entwicklungen, *within-person*, stehen für die Entwicklungen einer Person, interindividuelle Entwicklungen, *between-person*, für die unterschiedlicher Personen. Auch individuelle Gruppenentwicklungsmuster können analysiert werden. Exogene Einflüsse, welche die Wachstumsfaktoren beeinflussen, können in die Modelle mit aufgenommen werden. Der Vorteil liegt in der simultanen Schätzung der freien Parameter, welche im Strukturmodell Pfadkoeffizienten, Kovarianzen, Varianzen und Mittelwerte sind, im Messmodell sind dies Faktorladungen, Kovarianzen, Varianzen und Mittelwerte.

Als Schätzverfahren wird meist *maximum likelihood* (ML) verwendet. Dies setzt kontinuierliche normalverteilte Items voraus. Für geordnete kategoriale Items steht *weighted least squares* (WLS) zur Verfügung, welches diese Annahmen nicht hat. Dabei wird davon ausgegangen, dass dem kategorialen Item eine normalverteilte kontinuierliche Variable zugrunde liegt, die direkt nicht gemessen werden kann. Problem dabei ist die hohe Anzahl an Datensätzen, 2000 bis 5000, die benötigt wird, um unverzerrte Schätzwerte für Standardfehler und χ^2 -Werte zu erhalten. Die Parameter können zwar relativ stabil geschätzt werden, aber beispielsweise der χ^2 -Anpassungstest und andere χ^2 -basierte Fit-Indizes können verzerrt sein. Als Lösung bietet MPlus den *weighted least square estimator with variance and mean adjusted chi-square statistic* (WLSMV), welcher auch bei 150 bis 200 Datensätzen relativ stabile Schätzungen bringt. Zuerst werden Schwellenwerte bzw. *thresholds* geschätzt, dann die latenten Korrelationen, welche polyserielle Korrelationen zwischen den beobachteten Items und den angenommenen zugrundeliegenden Variablen darstellen. Schließlich werden die polychorischen Korrelationen zwischen den angenommenen zugrundeliegenden Variablen berechnet. Dies ergibt anstatt der Kovarianzmatrix aus der klassischen SEM-Schätzung eine Korrelationsmatrix. Für diese wird ein Schätzer für die Kovarianzmatrix der latenten Korrelationen erzeugt und für die WLS-Schätzung der freien Parameter verwendet. Zu beachten ist hier, dass eine normalverteilte Variable zu Grunde liegen muss. Dies ist aber möglicherweise nicht gerechtfertigt. Der Fehler, wenn rein monotone Beziehungen zwischen kategorialen Items zur Berechnung der linearen Korrelationen angenommen werden, ist aber als schwerwiegender zu betrachten. Deshalb ist eine mehrstufige Schätzung möglich, bei welcher zuerst die

ordinalen Wahrscheinlichkeiten für die Fälle berechnet werden. Dann wird eine latente Skala mit Normalverteilung erstellt und schließlich das Wachstumsmodell der latenten Variablen geschätzt.

Beim WLSMV- χ^2 -Test werden die im Modell errechneten Werte über die polychorischen Korrelationen mit den empirischen verglichen. Bei kleinen Fallzahlen ist die Anpassung möglicherweise verzerrt und mit Vorsicht zu genießen. Weitere Indizes wie *Comparative Fit Index* (CFI>0,95), der *Tucker Lee Index* (TLI>0,95), der *Root Mean Square Error of Approximation* (RMSEA<=0,06) oder der *Weighted Root Mean Square Residual* (WRMR<0,90) werden verwendet. Die ersten drei sind χ^2 -basiert, die letzten beiden nicht. Zum Modellvergleich dürfen χ^2 -Statistik Werte nicht verwendet werden, da die Differenz zweier robust geschätzter χ^2 -Werte nicht χ^2 -verteilt ist (Urban, 2004). Ist der χ^2 -Wert signifikant, bedeutet dies, dass das Modell nicht zu den beobachteten Daten passt (Acock und Li, 1996). RMSEA misst den *misfit* in der Population und nicht in der Stichprobe. Auch ein Konfidenzintervall zur Präzisionsmessung kann berechnet werden. Da χ^2 -verteilte Tests bei Verletzung von Verteilungsannahmen und hoher Beobachtungsanzahl nicht tauglich sind, empfiehlt sich der *Standardized Root Mean Square Residual* (SRMR), das Quadrat der absoluten Differenz zwischen beobachteten und modellbedingten Korrelationen. Der *Nonnormed Fit Index* (NNFI) vergleicht das Modell mit dem Nullmodell. Der Modellvergleich zwischen genesteten und nicht genesteten Modellen stellt eine weitere Möglichkeit der Modellevaluation dar. Für nicht genestete Modelle bieten sich AIC und BIC an.

Der typische Modellvergleich beginnt beim Null-Modell ohne Veränderung und lediglich einem *intercept*-Mittelwert mit fixierter Varianz null. Darauf folgt das *random intercept*-Modell bei welchem zusätzlich die *intercept*-Varianz geschätzt wird. Die nicht erklärte Varianz wird somit zwischen *intercept*-Varianz, auch interindividuelle Varianz genannt, und Störvarianz, auch intraindividuelle Varianz genannt, aufgeteilt. Sind diese signifikant, so ist die nicht erklärte Varianz groß genug um ein Modell mit mehr Parametern zu testen, das *fixed intercept fixed slope*-Modell. Dabei werden jeder Person gleicher Startwert und Steigung unterstellt. Es gibt keine interindividuelle Variation. Darauf folgt das *random intercept fixed slope*-Modell, welches interindividuelle Variation durch die *intercept*-Varianz erlaubt. Ist die Störvarianz groß genug, kann ein *random intercept random slope*-Modell getestet werden, wobei *slope*- und *intercept*-Varianz zufallsbedingt geschätzt werden. Somit ist eine Kovarianz zwischen beiden Varianzen möglich, welche bei Signifikanz interpretierbar ist. Es muss angemerkt werden, dass keine separaten *intercepts* und *slopes* geschätzt werden,

sondern eine multivariate Normalverteilung der latenten Variablen unterstellt wird, wodurch Mittelwerte, Varianzen und Kovarianzen über diese Verteilungen geschätzt werden. Die Modelle werden über einen χ^2 -Differenztest verglichen.

Zugrundeliegende Gruppen können als Modellerweiterung eingeführt werden. Dafür wird die Stichprobe in zwei Gruppen aufgeteilt und die Parameter simultan geschätzt. Werden Unterschiede zwischen zwei Gruppen vermutet, wird das Null-Modell ohne Einschränkungen hinsichtlich *intercept*- oder *slope*-Mittelwert oder Kovarianz, mit dem Modell mit Einschränkung über einen χ^2 -Differenztest verglichen. Die Gruppeninformation kann auch über einen externen Prädiktors, eine auf Level 2 zeitinvariante Kovariate, eingebracht werden. Diese misst die *between-individual*-Varianz. Das Modell wird vom *unconditional* zum *conditional* Modell. Bei einer Beeinflussung des *slopes* wird dies *cross level interaction* genannt, da die Zeit (Level 1) mit der Kovariate (Level 2) verknüpft wird. So kann auch eine zeitvariante mit einer zeitinvarianten Kovariate verbunden werden. Auch *latent growth mixture models* (LGMM) berechnen gruppenspezifische Trajektorien und werden später anhand der Kinodaten auch gezeigt. Dabei werden nicht beobachtete latente Gruppen mitgeschätzt, welche unterschiedliche in sich homogene aber zwischen den Gruppen heterogene Wachstumsverläufe aufweisen.

Mit Hilfe eines *parallel process*-Modells können Einflüsse paralleler Entwicklungen aufeinander gemessen werden, indem Kovarianzen unterschiedlicher *intercepts* und *slopes* erlaubt werden. Diese werden *multivariate change model*, *cross-domain individual growth model*, *multiple-domain model*, *fully multivariate latent trajectory model*, *simultaneous growth model*, *bivariate growth model* oder *associative latent growth model* genannt. Auch kausal gerichtete Einflüsse über Pfade zwischen Wachstumsfaktoren sind denkbar. Beispielsweise wird der *intercept*- auf den *slope*-Faktor regressiert (Preacher et al., 2008) oder die Wachstumsfaktoren beschreiben einen *outcome*.

Reddy et al. (2003) bringen ein *cross-domain latent growth modeling*-Beispiel. Studenten werden in drei aufeinanderfolgenden Jahren nach empfunderer Unterstützung seitens der Lehrer (Kurzversion der Classroom Environment Scale), Selbstwertgefühl (*General Self-Esteem* Kurzversion des *Self-Evaluation Questionnaire*) und depressiven Symptomen (Kurzversion der *Children's Depressive Inventory*) befragt. Das Geschlecht wird gesondert betrachtet. Das Messmodell beschreibt die individuellen Entwicklungen der *outcome* Variablen (Depression und Selbstwertgefühl) und der Unterstützung. Das Strukturmodell beschreibt den Zusammenhang zwischen den latenten Wachstumsmodellen,

outcome Variable und Unterstützung. Heterogenität wird im linearen Modell durch Varianzen von *intercept* und *slope* aufgefangen. Auf Level eins wird das *within-person*-Modell definiert. Die Verbindung dieser auf Level zwei wird *between-person*-Modell genannt. Zur Schätzung wird *full information maximum likelihood* (FIML) herangezogen, welcher den EM Algorithmus nutzt um Mittelwerte und Kovarianzen zu schätzen. Daraus werden die Startwerte für die ML-Prozedur berechnet. FIML nützt alle Werte und liefert die *maximum likelihood chi square* und den RMSEA. Zuerst werden drei *unconditional*-Modelle geschätzt welche signifikante Mittelwerte und Varianzen für *intercept* und *slope* aufweisen. Im Anschluss werden diese im kombinierten Modell zusammengehängt, wobei Unterstützung die Depression und das Selbstwertgefühl beeinflusst. Um den Geschlechterunterschied zu testen, wird das Geschlecht als Moderator-Variable im Sinne einer *multigroup*-Analyse eingeführt. Zuerst wird frei geschätzt, dann werden die Pfade für Mädchen und Buben gleichgesetzt und auch bestätigt. Residuenvarianzen und Kovarianzen für die beiden *outcome*-Variablen sind nicht geschlechterunterschiedlich, die Unterstützung schon. Ein Test auf unterschiedliche *intercepts* und Levels in den drei Konstrukten zeigt für alle drei Konstrukte unterschiedliche Modellbevorzugungen. Das Geschlecht als Moderatorvariable des Unterstützungseinflusses auf die beiden *outcome*-Konstrukte wird nicht bestätigt. Vergleiche zwischen Gruppen, männlich vs. weiblich und vier künstlich gebildeter Gruppen (Gruppe 1: hoher Startwert - eine Standardabweichung über dem Mittelwert und steigende Unterstützung, Gruppe 2: hoher Startwert und progressiv sinkende Unterstützung, Gruppe 3: niedriger Startwert - eine Standardabweichung unter Mittelwert und steigende Unterstützung, Gruppe 4: niedriger Startwert und sinkende Unterstützung) bezogen auf die *outcome*-Variablen zeigen Unterschiede. Für die umgekehrte Beeinflussung, wobei das Selbstwertgefühl und Depressionen die empfundene Unterstützung beeinflussen, werden die Richtungen umgedreht. Das Modell hat einen schlechten Fit, beinahe keine signifikanten Pfade und wird verworfen. Einschränkungen der Studie betreffen die gleichzeitige Beobachtung und die daraus folgende eingeschränkte kausale Aussagekraft.

Bei mehrmaliger Befragung und einhergehender Panelmortalität bieten *cohort-sequential designs*, auch *accelerated longitudinal designs*, eine Lösung. Angenommen, Kohorte eins wurde in den Jahren 1999, 2000, 2001 und 2002 befragt, Kohorte zwei in den Jahren 2001, 2002, 2003 und 2004. Um einen Gesamttrend zu messen bieten sich zwei Möglichkeiten. In der ersten werden die beiden Datensätze zusammengelegt und die fehlenden Werte als MCAR behandelt, da sie aufgrund des Befragungsdesigns fehlen. In der zweiten werden

die Kohorten als separate Gruppen behandelt, wobei die Parameter für die beiden Gruppen gleich gesetzt werden. Die Summe aus *intercept* und *slope* von Kohorte eins des letzten Beobachtungszeitpunktes im Jahre 2002, wird mit der zweiten Beobachtung von Kohorte zwei gleichgesetzt, etc. In Modellen mit zeitvarianten Kovariaten werden diese als wiederholte Prädiktoren eingebracht. Der Einfluss kann sich zeitlich ändern, über die Individuen gleich sein oder auch konstant gehalten werden. Auch unterschiedliche Einflüsse für unterschiedliche Individuen können geschätzt werden.

Im polynomialen Wachstumsmodell werden quadratische Effekte über einen zusätzlichen Wachstumsparameter, den quadratischen *slope*-Faktor, geschätzt. Die Sinnhaftigkeit dieses Terms, verglichen mit dem sparsamen linearen Modell, wird über einen χ^2 -Differenztest beurteilt und auch für das Kinobeispiel gezeigt. Nicht-lineare Veränderungen können auch mit Hilfe von *unspecified trajectories*-Modellen geschätzt werden, auch *completely latent* oder *fully latent* genannt. Dabei werden die Faktorenladungen des Steigungsfaktors frei geschätzt. Die erste Ladung wird auf null fixiert, die Ladung der zweiten Beobachtung oder in Spezialfällen auch eine der folgenden, meist auf eins. So wird im ersten Fall zwischen den ersten beiden Beobachtungen die Größe der Veränderung geschätzt. Die frei geschätzten Ladungen der folgenden Beobachtungen geben den Multiplikator für den Veränderungswert der folgenden Beobachtungen an. Auch dieses Modell wird hier empirisch vorgestellt. Werden einzelne Ladungen im linearen Modell frei geschätzt, kann ein χ^2 -Differenztest die Abweichung vom linearen Verlauf zwischen den genesteten Modellen prüfen. Auch *structured latent curve* (SLC)-Modelle, in welchen die Ladungsmatrix Werte einer zugrundeliegenden Funktion darstellen, sind denkbar. *Piecewise growth models* oder auch *discontinuity design* werden hier später ebenfalls verwendet, um sich ändernde Verläufe zu spezifizieren. Dabei steigt der Verlauf in den ersten Beobachtungen beispielsweise nicht-linear und fällt später linear ab. Dafür werden zwei *slopes* geschätzt, welche sich überlappen, oder das Ende des Verlaufes mit dem ersten *slope* bildet den Beginn des Verlaufes mit dem zweiten *slope*.

Der Vorteil der SEM-Lösung liegt in der Untersuchung der Veränderung latenter Konstrukte mit mehreren Indikatoren. Die *common variance* wird von der *unique variance* getrennt, sprich, die Messfehlervariation wird von der Variation um den Wachstumsmittelwert getrennt. Die Modelle werden *second order growth curve models*, *curve-of-factors models*, *latent variable longitudinal curve models* oder *second-order latent growth models* genannt. Die Faktoren erster Ordnung kommen durch die Verbindung mehrerer Indikatoren zustande. Die

Wachstumsfaktoren bilden die Faktoren zweiter Ordnung. Um die inhaltliche Stabilität der Faktoren zu gewährleisten, muss die Stationarität bzw. Invarianz der Faktorladungen untersucht werden. Wie zuvor erwähnt gibt es einen Gegensatz zu den Ein-Indikator-Modellen. Die Varianz in Messmodellen mit mehreren Items, welche durch einen Faktor verursacht wird, soll möglichst groß sein und die Messfehlervarianz abgetrennt werden. Die Items repräsentieren schließlich nicht vollständig das latente Konstrukt, sondern in jener Stärke, wie sie durch die Faktorladungen ausgedrückt werden. Diese wird *true score common variance* genannt und wirkt sich auch auf die Korrelationen zwischen den Items aus. Je höher die Faktorladungen, desto besser die gemeinsame Faktorstruktur und umso kleiner der Anteil der Residualvarianz an der Gesamtvarianz.

4.3.6.1. Modellspezifikation

Der beobachtete Wert einer Person für ein Item zu einem bestimmten Zeitpunkt, y , wird linear additiv von einer indicatorspezifischen Konstante, *intercept*, für jedes Item zu einem bestimmten Zeitpunkt, v , den Einflüssen der Faktoren, nämlich der Faktorladung mit dem ein Item zu einem bestimmten Zeitpunkt beeinflusst wird, λ , dem Faktorenwert zu einem bestimmten Zeitpunkt, η , und einem Residualwert für jedes Item zu jedem Zeitpunkt, welcher durch Messfehler oder Störeffekte entsteht, über alle Personen gleich null ist und zu keinem Zeitpunkt mit einem Faktor korreliert, ε , bestimmt. Daraus folgt Formel 19, auch *data-Modell* genannt (Preacher et al., 2008).

$$y_{ij} = v_j + \lambda_{jk}\eta_{ik} + \dots + \lambda_{jm}\eta_{im} + \varepsilon_{ij}$$

Formel 19: Messmodell (Urban, 2004)

Das Strukturmodell im linearen Fall ist in Formel 20 dargestellt. η ist der Faktorenwert der Person zu einem gewissen Zeitpunkt. Die weiteren η bestimmen die Werte des Levelfaktors und des Steigungsfaktors z sind die Zeitkoeffizienten mit fixierten Werten. ζ ist der Residualanteil des Faktorenwertes einer Person zu einem bestimmten Zeitpunkt. Dieser hat zu jedem Zeitpunkt den Erwartungswert null. Die Residualwerte korrelieren zu keinem Zeitpunkt mit dem Faktor.

$$\eta_{it} = \eta_{i1}z_{1t} + \eta_{i2}z_{2t} + \zeta_{it}$$

Formel 20: Strukturmodell (Urban, 2004)

Preacher et al. (2008) formulieren die Bestimmung der Faktoren in Formel 21. Die Notation weicht von jener von Urban (2006) ab. η_{1t} und η_{2t} sind die

random coefficients oder *random effects*. ζ_{1i} und ζ_{2i} geben die individuellen Abweichungen, also die Residuen von den latenten Mittelwerten α_1 und α_2 an.

$$\begin{aligned}\eta_{1i} &= \alpha_1 + \zeta_{1i} \\ \eta_{2i} &= \alpha_2 + \zeta_{2i}\end{aligned}$$

Formel 21: Latente Variablen (Preacher et al., 2008)

Wird nur ein beobachtetes Item analysiert und nicht wie in obigen beiden Formeln dazwischen eine konfirmatorische Faktorenanalyse eingelesen, wird dessen Wert wie in Formel 22 bestimmt. y ist der Itemwert einer Person zu einem gewissen Zeitpunkt. η ist der Wert des Wachstumsfaktors. z sind die Zeitkoeffizienten. ε ist der Residualanteil der Person zum jeweiligen Zeitpunkt. Der Residualanteil hat wiederum über alle Personen zu jedem Zeitpunkt den Erwartungswert null. Die Residualwerte korrelieren nicht mit dem Faktor.

$$y_{it} = \eta_{1i}z_{1t} + \eta_{2i}z_{2t} + \dots + \eta_{mi}z_{mt} + \varepsilon_{it}$$

Formel 22: Ein-Item-Modell (Urban, 2004)

Zur Identifikation der Messmodelle muss jeweils eine Ladung eines Faktors auf eins fixiert werden. Die weiteren Faktorladungen der Messmodelle sollten eine annähernd stabile Form über die Zeit aufweisen. Verändert sich diese zu stark, bedeutet dies, dass die latenten Konstrukte über die Zeit einen variierenden Inhalt beschreiben. Diese Veränderung steht konträr zu der mit Hilfe der Wachstumsmodelle gewünschten Untersuchung, ob sich ein bestimmtes Konstrukt ändert. Deshalb werden die Messmodelle modifiziert und invariant über die Zeit spezifiziert. Bei der Behandlung der Items als kategoriale Items, z.B. im Zuge des WLSMV-Schätzalgorithmus, ist weiters darauf zu achten, dass auch die Schwellenwerte über die Zeit konstant gehalten werden. Sonst kann nicht gewährleistet werden, dass die Veränderungen auf inhaltlichen und nicht auf Veränderungen seitens der Messskala basieren. Deshalb werden im Zuge von Wachstumsmodellen meist indikatorspezifische Schwellenwert-Invarianzen geschätzt.

Zusätzlichen Modellfit bringt die freie Schätzung indikatorspezifischer und zeitpunktübergreifender Korrelationen zwischen den Residualvarianzen, die diachronen Korrelationen. Somit korrelieren die Residualvarianzen als auch die Messwerte der Items. Synchrone Korrelationen, also jene zwischen unterschiedlichen Items zu einem Zeitpunkt sind nicht indikatorspezifisch und nicht gebräuchlich. Ebenfalls nicht eingeführt werden Korrelationen zwischen unterschiedlichen Items zu unterschiedlichen Zeitpunkten, welche zwar

diachron, also zeitübergreifend wären, aber nicht indikatorspezifisch. „Uns sind keine empirischen Längsschnittanalysen bekannt, die bei Verwendung von Mehr-Indikatoren-Messmodellen ohne diese Form von Fehlerkorrelationen akzeptable Fitwerte erreicht hätten“ (Urban, 2004). Begründet wird dies über die Gesamtvarianz. Diese unterteilt sich in die *common variance*, also die gemeinsame oder wahre Varianz und in die Residualvarianz, welche aufgrund von zufälligen und systematischen Messfehlern entsteht. Die systematische, also spezifische bzw. *true score unique*-Varianz der Items ist keine Fehlervarianz, sondern sie entsteht aufgrund der Beeinflussung der Items durch andere Variablen oder Faktoren, welche nicht im Modell enthalten sind und nicht über indikatorspezifische Kovarianzen einfließen. Lediglich Korrelationen zwischen Residuen zeitlich benachbarter Items sind empfehlenswert, welche aufgrund von Erinnerungseffekten auftauchen. Um das Modell durch derartig begründete Annahmen nicht einzuengen, empfiehlt sich oben genannte Spezifikation, welche zeitunabhängig ist und Verursacher aufgrund von Erinnerungseffekten, als auch andere Gründe beinhaltet. Durch die Einführung der diachronen Residuenkorrelationen sinken die Faktorkorrelationen, da systematische Varianzen nun durch diese aufgenommen werden. Auch eine *first-order autoregressive* Fehlerstruktur ist denkbar. Dabei werden die Kovarianzen benachbarter Fehlerterme frei geschätzt. Auch eine kausale Fehlerkovariation ist denkbar (Preacher et al., 2008).

Die Residuenvarianzen und -kovarianzen repräsentieren jenen Teil der Varianz, welcher durch die latenten Kurven nicht aufgefangen wird. Homoskedastische Residuen werden durch eine Gleichheitsbeschränkung der Kovarianzen spezifiziert (Preacher et al., 2008). Korrelationen zwischen Residuen müssen inhaltlich theoriebedingt begründet werden, da sie keine Fehlerwerte im herkömmlichen Sinne darstellen. Sie können Varianzanteile beinhalten, welche durch externe Prädiktoren erklärt werden könnten.

Die Ladungsstruktur eines Faktors, dessen Veränderung über die Zeit hinweg untersucht wird, darf sich nicht verändern, da unterschiedliche Ladungen für unterschiedliche Itemeinflüsse stehen. Die robusten bzw. zeitinvarianten Faktorstrukturen können in folgender unterschiedlicher Einschränkungstärke definiert werden (Urban, 2004). Striktere Spezifikationen werden nach Rechtfertigung über den χ^2 -Test oder über logische Begründungen eingeführt.

- 1) Konfigurative Faktorinvarianz: relative, aber nicht absolute, zeitlich gleich bleibende Ladungsgrößen, somit auch zeitlich ordinal gleiche Anordnung der Ladungen;

- 2) Metrische (metrisch-partielle, metrisch-saliente) Faktorinvarianz: gleiche Ladungsgrößen gleicher Items über die Zeit;
- 3) Skalare (bzw. strikt-skalare, partiell-skalare) Faktorinvarianz: latente Mittelwerte der Faktoren gleich, macht keinen Sinn bei Längsschnittanalysen, da genau diese Veränderung untersucht wird; in SEM-basierten LGC-Lösungen zwingt diese Einschränkung jedoch die zeitvarianten Eigenschaften der zeitinvariant betrachteten Faktoren auf die latente Ebene;
- 4) Varianzbezogene Faktorinvarianz: alle latenten Faktorvarianzen werden auf eins fixiert, nicht sinnvoll bei polarisierenden Entwicklungen da diese dann nicht gemessen werden können;
- 5) Strikte Faktorinvarianz: Kombination aller genannten Restriktionen führt zu identischen Ladungsmatrizen, Faktorenmittelwerten und Faktorenvarianzen für jedes Messmodell, dient maximal für Gruppenvergleiche, ist aber für Längsschnittanalysen nicht tauglich;

Sowohl intraindividuelle Entwicklungsmuster, als auch gruppentypische Entwicklungsmuster wie unterschiedliche Startwerte oder Veränderungsraten können analysiert werden. Unterschiedliche Formen der Veränderung wie lineare, kurvilineare, oder zeitlich begrenzte Entwicklungen, als auch zeitinvariante oder zeitvariante Prädiktoren bzw. Kovariate können das Ausgangsniveau als auch die Veränderungsrate beschreiben und beeinflussen. Das hier verwendete SEM-basierte LGC-Modell, welches Wachstumsprozesse von latenten Faktoren untersucht, wird *curve-of-factors*-Modell genannt und beschreibt *multivariate* LGC-Modelle mit mehreren Indikatoren. Somit werden Modelle zweiter Ordnung realisiert. Derartige Modelle können auch über Mehrebenenmodelle gelöst werden. Urban (2004) beschreibt fünf Annahmen, welche hinter den LGC-Modellen stecken.

- 1) Für jede Person gibt es einen wahren zugrundeliegenden personenspezifischen Entwicklungspfad.
- 2) Der personenspezifische Pfad ist nicht identisch mit den beobachteten Werten, da letztere durch Störeinflüsse verzerrt werden.
- 3) Der wahre Entwicklungspfad kann aufgrund mehrerer beobachteter Items bestimmt werden.
- 4) Alle Individuen einer Population unterliegen gemeinsamer interindividueller Basisfunktionen, welche die Verlaufsform bestimmen.
- 5) Da die Individuen die gemeinsamen Basisfunktionen unterschiedlich nutzen, können diese individuell auch bei verschiedenen Startwerten starten und unterschiedliche Veränderungsraten annehmen. Die Verlaufsform bleibt aufgrund der vierten Annahme jedoch für alle gleich.

Die Nutzung oder Gewichtung variiert zufällig über die Population und wird derart geschätzt, dass die Differenz der im Modell erzeugten und empirisch beobachteten Werte, Kovarianzen und Mittelwerte der Daten, ein Minimum werden lässt. Ein schlechter Fit bedeutet also möglicherweise eine falsche Spezifikation der zugrunde liegenden Wachstumskurve oder ein Nicht-Vorhandensein dieser. Die Heterogenität zwischen den Individuen kann auch zu groß sein und muss dann durch externe Prädiktoren aufgefangen werden. Auch unpassende Residuenkovarianzen, spezifizierte konstante Schwellenwerte bei kategorialen Variablen, welche sich eigentlich ändern, bzw. stabile Faktorstrukturen, welche verändernde Faktorladungen aufweisen, können den Modellfit negativ beeinflussen. Als Fitbeurteilung muss somit der Anteil der erklärten Varianz berücksichtigt werden. Ist der Faktorenmittelwert nicht signifikant, liegt möglicherweise keine Basisfunktion über alle Personen vor. Urban (2004) merkt an, „dass z.B. der Gesamtfitt zur Beurteilung von LGC-Modellen überhaupt keine Rolle spielt, wenn diese als eine Variante der Mehrebenenanalyse im Kontext der dort üblichen Berechnungsverfahren modelliert und geschätzt werden“.

Um die Entwicklungsmuster zu bestimmen muss eine Basisfunktion mit Wachstumsfaktoren das Ausgangslevel und die weitere Verlaufsform bestimmen. Die Varianzen der latenten Wachstumsfaktoren bestimmen die interindividuellen Differenzen und die Mittelwerte den interindividuellen Verlauf. Die Wachstumsfaktoren werden mit den latenten Konstrukten der Messmodelle verbunden. Die Ladungen werden auch Zeitkoeffizienten genannt. Das Ausgangslevel weist über alle Zeitpunkte die gleichen Ladungen auf und geht zu jedem Zeitpunkt in die Berechnung mit ein. Für den ersten Zeitpunkt gibt es noch keine Veränderung. Somit bildet der Level über alle Zeitpunkte auch gleichzeitig den Startwert. Die Ladungen der weiteren Wachstumsfaktoren bestimmen die Form der Veränderung. Für einen linearen Verlauf genügt ein Wachstumsfaktor, für einen quadratischen Verlauf wird ein weiterer Wachstumsfaktor benötigt. Die Ladungen des Wachstumsfaktors für den linearen Einfluss nehmen in zeitlich konstanten Zeitabschnitten die Werte 0, 1, 2, 3, 4, 5, etc. an. Um einen linearen Verlauf bestätigen oder widerlegen zu können, müssen mehr als zwei Zeitpunkte beobachtet werden, sonst gibt es keine Freiheitsgrade (Acock und Li, 1996). Um die monatlichen Veränderungen zu messen müssen diese auf 0, 12, 24, 36, 48, 60, etc. fixiert werden. Weisen die Beobachtungszeitpunkte unterschiedliche Intervalle auf, dient diese Lösung beispielsweise zur Spezifizierung eines linearen Verlaufes. Wurde im Jahr 1996, 1998, 1999, 2005, und 2006 gemessen, wird ein jährlich linearer Verlauf mit den Zeitkoeffizienten 0, 2, 3, 9, 10 bestimmt. Je nachdem welcher Zeitpunkt als

Hauptzeitpunkt gesehen wird, kann die Fixierung einer Ladung auch auf einen anderen Zeitpunkt verlegt werden. Beispielsweise sind die Ladungen -5, -4, -3, -2, -1, 0 auf den letzten Zeitpunkt zentriert und müssen auch dahingehend interpretiert werden. Die *slope*-bezogenen Parameter, wie beispielsweise *slope mean* oder den *slope* beeinflussende Regressionsparameter, verändern sich. Der Modellfit bleibt gleich. Die Fixierung sollte jenen Zeitpunkt betreffen, welcher um den *intercept* zu interpretieren am wichtigsten erscheint (Preacher et al., 2008). Es wird empfohlen den Zeitpunkt mit der geringsten Streuung zu wählen.

Auch nicht-lineare vorweg bestimmte Verläufe können so definiert werden, beispielsweise eine Steigung bis Zeitpunkt drei und danach ein sinkender oder gleichbleibender Verlauf. Die Ladungen des quadratischen Wachstumsfaktors nehmen in zeitlich konstanten Zeitabschnitten die Werte 0, 1, 4, 9, 16, 25, etc. an. Erhält der zusätzliche Term ein negatives Vorzeichen, verläuft die Entwicklung konkav, also U-förmig zur Zeitachse, ist er positiv verläuft sie konvex, also umgekehrt U-förmig zur Zeitachse. Durch spezielle Bestimmung der Zeitkoeffizienten lassen sich auch mit Hilfe von nur zwei Wachstumsfaktoren beispielsweise logarithmische oder exponentielle Verläufe definieren. Um den tatsächlichen Wert für einen Zeitpunkt berechnen zu können, vorausgesetzt es handelt sich um eine Skala mit original interpretierbaren Werten, wird zum Levelmittelwert das Produkt aus Zeitkoeffizient multipliziert mit dem Mittelwert des Steigungsfaktors addiert. Liegt der Wert der *t*-Statistik $t(\text{var}_F) = \text{var}_F / \text{std.err}_F$ für die Mittelwerte oder die Varianzen über 1,96 für den zweiseitigen Test, gibt es signifikante Mittelwerte für den Level- und den Steigungsfaktor bzw. weist die beobachtete Stichprobe signifikant hohe Heterogenität auf und es existiert keine zugrunde liegende Wachstumsfunktion. Auch die Korrelation zwischen Level- und Steigungsfaktor kann interpretiert werden. Ist die Varianz des Steigungsfaktors nicht signifikant, wird die Kovarianz zwischen Level- und Steigungsfaktor auch nicht signifikant sein. Dies bedeutet, dass es keinen Zusammenhang zwischen der Höhe des Startwertes und der Entwicklung gibt. Je niedriger die Korrelation, desto stärker die externen Einflüsse, welche diese Beziehung beschreiben. In der vorliegenden Kinostudie kommt es zu negativen Korrelationen, welche später interpretiert werden. Um zu überprüfen ob eine Steigung existiert, wird das Modell mit Steigung mit dem rein latenten Mittelwert-Modell ohne Steigung, in welchem der Steigungsfaktor auf null fixiert wird, verglichen. Besteht laut χ^2 -Test kein signifikanter Unterschied, wird das komplexere Modell mit Steigungsfaktor zu Gunsten des einfacheren Mittelwertmodells verworfen.

Da die Varianzen individuelle Trajektorien auffangen, bestimmt deren Höhe die interindividuellen Differenzen. Veränderungen der Varianzen geben Auskunft über die Zunahme oder Abnahme an Heterogenität zwischen den Individuen. Für gewöhnlich wird von Null-Varianz ausgegangen. Existieren signifikante Unterschiede in den Varianzen, kann der Versuch unternommen werden, diese im Modell zu erklären. Das Modell besteht folglich aus einer *confirmatory factor analysis* (CFA) mit fixierten Zeitkoeffizienten der Wachstumsfaktoren.

Um kurvilineare Veränderungen schätzbar zu machen, müssen meist mindestens vier Zeitpunkte erhoben werden, sonst ist das Modell unteridentifiziert bzw. nicht sinnvoll interpretierbar. Je mehr Zeitpunkte desto besser. Eine Ausnahme bietet die freie Schätzung der Wachstumsfunktion. Um beispielsweise ein lineares Wachstum über drei Zeitpunkte zu bestimmen, müssen mindestens acht freie Parameter geschätzt werden, nämlich Mittelwert der Steigungsfaktoren, der Mittelwert des Levelfaktors kann auf null fixiert werden, die Varianzen des Level- und des Steigungsfaktors und die Kovarianz zwischen diesen, und drei Residuenkovarianzen, welche durch andere Spezifizierungen erweiterbar sind. Wird nur ein Item zu den drei Zeitpunkten beobachtet, so stehen nur $p(p+1)/2$ Parameter zur Verfügung. p steht für die Anzahl an beobachteten Items. Mit nur einer manifesten Variablen ist das Modell unterspezifiziert und folglich nicht schätzbar.

Auch Modelle mit nicht-linearem Verlauf, welcher nicht vordefiniert ist, können modelliert werden. Bei drei Zeitpunkten und frei geschätzten Zeitkoeffizienten muss ein weiterer Parameter zur Schätzung freigegeben werden, der Zeitkoeffizient zum Zeitpunkt drei. Zum Zeitpunkt eins gibt es noch keine Veränderung. Zum Zeitpunkt zwei wird der Mittelwert des Steigungsfaktors bestimmt, dieser steht für die Veränderung zwischen Zeitpunkt eins und zwei. Wird die Veränderung zwischen Zeitpunkt zwei und drei aus der Datenstruktur explorativ ermittelt, wird der Zeitkoeffizient nicht auf zwei wie im linearen Modell fixiert, sondern frei geschätzt. Er gibt an, um das wievielfache sich die Steigung zwischen Zeitpunkt zwei und drei gegenüber der Steigung zwischen Zeitpunkt eins und zwei verändert hat. Ist der Zeitkoeffizient größer als zwei, bedeutet dies eine Steigerung der Steigung, ist er kleiner sinkt die Veränderung im zweiten Zeitintervall. Anhand der Differenz zwischen linearen und frei geschätzten Zeitkoeffizienten wird der Unterschied zwischen linearer und nicht-linearer Veränderung einsehbar gemacht. Ein nicht-linearer Verlauf bestätigt sich, wenn der Großteil der individuellen Trajektorien deutlich oberhalb oder unterhalb des linearen Verlaufes liegt. Ist das komplexere frei geschätzte Modell nicht signifikant unterschiedlich vom linearen Modell mit einem zusätzlichen

Freiheitsgrad, wird das frei geschätzte Modell verworfen. Diese Interpretationen setzen voraus, dass die Steigung zum Zeitpunkt null auf null fixiert wird und der Zeitkoeffizient zum Zeitpunkt eins auf eins. Alternativ kann auch ein anderer Zeitkoeffizient auf eins fixiert werden, erschwert jedoch die Interpretation der Ergebnisse. Die Zentrierung der Modelle bestimmt nämlich jenen Zeitpunkt, wo der Zeitkoeffizient auf null fixiert wird. Wird der erste Zeitkoeffizient auf null fixiert, wird das Modell *initial status*-Modell genannt, wird der mittlere auf null fixiert, wird dies zeitdurchschnittliches Modell genannt. Die Anpassungsqualität, also der Modellfit über den χ^2 -Wert, als auch die Residualvarianzen, werden von der Zentrierung nicht beeinflusst. Lediglich der Levelfaktor und die Korrelation zwischen Level- und Steigungsfaktor variiert. Der latente Mittelwert und die Varianz des Steigungsfaktors bleiben ebenfalls gleich.

Sind die latenten Varianzen signifikant, kann nicht von einem Varianzwert von null ausgegangen werden. Ist der Steigungsfaktor signifikant, deutet dies auf gruppentypisches Wachstum mit individuellen Unterschieden hin. Gruppierungen sind möglich, welche durch endogene Prädiktoren einen Anteil der nicht erklärten Wachstumsfaktorvarianzen erklären. Derartige zeitinvariante oder zeitvariante Prädiktoren werden in *conditional* Modellen eingebracht. Um die Relevanz eines Prädiktors zu überprüfen, wird der proportionale Rückgang der Varianz berechnet. Die Differenz zwischen *conditional* Varianz (partielle Residualvarianz) und *unconditional* Varianz (Residualvarianz des Wachstumsfaktors) wird durch die *unconditional* Varianz dividiert und beschreibt den prozentuellen Rückgang erklärter Varianz. Für den zeitvarianten Effekt wird eine zusätzliche Modellgleichung aus Formel 23 benötigt.

$$\eta_{ji} = \alpha_j + w_{ki}\gamma_j + \zeta_{ji}$$

Formel 23: Konditionales LGC-Modell

η beschreiben die Wachstumsfaktoren für jede Person, α die *intercept*-Werte der Wachstumsfaktoren, w den Prädiktorwert der Person, γ die Koeffizienten der Prädiktoren und ζ die Residualanteile der Wachstumsfaktoren für die Personen und Faktoren. Somit kann auch die kausale Auswirkung des Levelfaktors auf den Steigungsfaktor betrachtet werden und Varianzanteile des Steigungsfaktors erklären. In Modellen mit zeitinvarianten Prädiktoren führen diese zu zeitvarianten Effekten, welche zu jedem Zeitpunkt anders auf die Items wirken. In rein zeitvarianten Modellen ist außer dem sich zeitlich verändernden Prädiktoreffekt auch der latente Prädiktor selbst aus zeitabhängigen latenten Prädiktorvariablen spezifiziert und es entstehen zeitvariante Effekte zeitvarianter Prädiktoren. Derartig umfangreiche Modellerweiterungen über Prädiktoren

werden häufig zur Stabilitätskontrolle der Wachstumsmodelle verwendet. Mit Hilfe von zeitinvarianten Kovariaten, beispielsweise Geschlecht, oder auch zeitvarianten Kovariaten, beispielsweise ein parallel beobachtetes Wachstum, können zusätzlich Varianzanteile erklärt werden. Auch zeitinvariante Kovariate, welche nur einen Zeitpunkt betreffen sind denkbar.

Urban (2004) beschreibt Modellerweiterungen. Beispielsweise kann ein universelles Wachstum durch ein indikatorspezifisches Wachstum erweitert werden. Es kann nämlich nicht immer davon ausgegangen werden, dass alle Items das gleiche Entwicklungsmuster aufweisen. Für jene die abweichen wird ein zusätzliches Wachstumsmodell mit Level- und Steigungsfaktoren hinzugefügt. Das *piecewise* LGC-Modell unterteilt Veränderungen in verschiedene Phasen. Für jede Phase wird ein eigenes Wachstumsmodell aufgesetzt. Der Endpunkt der ersten Phase ist gleichzeitig der Startpunkt der zweiten Phase, etc. Die unterschiedlichen Phasen können wiederum durch externe Prädiktoren beeinflusst werden. In *cross domain*-LGC-Modellen bzw. bereichsübergreifenden Modellen werden Kovarianzen zwischen Wachstumsfaktoren unterschiedlicher Wachstumsbereiche als Korrelationen oder Effektbeziehungen eingeführt. Diese können gerichtet rekursiv, gerichtet nonrekursiv oder auch ungerichtet, also Korrelationen sein. Verbindungen gleicher Art, zwischen Level- oder Steigungsfaktoren, oder auch ungleicher Art, zwischen Level- und Steigungsfaktoren, sind möglich. Sie können zeitlich parallel oder auch versetzt erfolgen. In *multigroup*-LGC-Modellen wird der Einfluss einer Intervention durch einen zusätzlichen Wachstumsfaktor mit der Kontrollgruppe verglichen, vorausgesetzt es besteht ausreichende Faktorinvarianz. Für einen Modellvergleich wird BIC verwendet, da die beiden Modelle eine unterschiedliche Faktorenanzahl besitzen. LGC *simplex*-Modelle, auch autoregressive Stabilitätsmodelle genannt, erweitern die klassischen LGC-Modelle um eine Untersuchung der autoregressiven Stabilität, also der Gleichförmigkeit von Entwicklungen. In autoregressiven oder *Markov simplex*-Modellen entsteht jede wiederholte Messung aus der vorhergehenden und einem verzerrten Anteil. Die Varianz der folgenden Beobachtung wird durch die vorangehende erklärt. Dieses Modell wird im nächsten Kapitel empirisch getestet. Dies führt zur Verknüpfung benachbarter Wiederholungen. Auch die Verarbeitung von Kohorten-Sequenz-Designs ist möglich. Dafür werden die kurzen Zeitreihen als Teile einer gesamten Zeitreihe betrachtet und eine durchgängige Wachstumsfunktion geschätzt. Der Levelfaktor ist für alle Gruppen gleich, die Zeitkoeffizienten für den Steigungsfaktor jedoch unterschiedlich. Die latenten Konstrukte müssen über die Gruppen hinweg als invariant spezifiziert werden. Auch Teilgruppen mit unterschiedlichen

Beobachtungshäufigkeiten lassen sich anhand von Fehlerwertdesigns oder *multiple group*-Analysen untersuchen, um durch die auftretende Panelmortalität nicht lediglich jene Personen analysieren zu können, welche zu jedem Zeitpunkt die Items bewertet haben, sondern eben nur über eine kürzere Zeitspanne. Dabei ist zu beachten, dass die Anzahl an unterschiedlichen Beobachtungsgruppen nicht zu groß sein darf. Gibt es mehrere Beobachtungssequenzgruppen, ist eine ML-Schätzung, welche der Annahme der multivariaten Normalverteilung folgt, nicht mehr tauglich. Anstatt der Mehrgruppenanalyse wird FIML verwendet. Er ist im Gegensatz zu Datenersatzmethoden, dem paar- oder listenweisen Löschen von Daten, effizienter, weniger verzerrt und konvergiert leichter wenn die Daten *missing completely at random* (MCAR) oder *missing at random* (MAR) sind (Rubin, 1976). Besser als paar- oder listenweises Löschen, was zu negativen Werten in der Kovarianzmatrix führen kann, sind beispielsweise *generalized least squares*, *unweighted least squares*, EM oder asymptotisch verteilungsfreie Methoden. FIML wird jedoch auch gegenüber diesen Methoden der Vorzug gegeben, da er die gesamte verfügbare Information nutzt (Preacher et al., 2008).

4.3.6.2. Schätzung von Modellen mit *missing values*

Enders und Bandalos (2001) vergleichen FIML mit listenweisem und paarweisem Fallausschluss und der *similar response pattern imputation*. Monte Carlo-simulierte Testdaten, welche zum Vergleich einer CFA als auch eines *full structural models* dienen, unterschieden sich in drei verschiedenen Dateneigenschaften, nämlich drei Faktorenladungsstärken, vier Stichprobengrößen und fünf unterschiedlichen Fehlerwertraten. Die Fehlerwertraten werden zufällig für MCAR-Bedingungen, für drei Variablen in Abhängigkeit sechs weiterer Variablen (jedoch nicht unter Abhängigkeit von sich selbst) und auch für MAR-Bedingungen, künstlich entfernt. Die Performance der Lösungen wird anhand vier unterschiedlicher Kriterien gemessen, dem Anteil an Konvergenzproblemen bei der dreifachen Anzahl an Iterationen gemessen an den zu schätzenden Parametern, am Schätzfehler der Parameter durch prozentuellen Vergleich des Stichprobenfehlers mit den wahren Parameterwerten, an der Schätzeffizienz der Parameter durch Vergleich der Stichprobenvarianz mit der geschätzten Varianz und anhand des Modellfits, gemessen anhand der prozentuellen Modellablehnung. Sämtliche Konstellationen führen zur Präferenz des FIML. FIML liefert die meisten brauchbaren Lösungen, wird hinsichtlich Modellfit bevorzugt und liefert unverzerrte und effizientere Schätzungen sowohl für MCAR- als auch MAR-Fehlerwertstrukturen. Listenweiser und paarweiser Fallausschluss liefern unverzerrte Schätzungen für MCAR, jedoch nur FIML auch für MAR. Die Ergebnisse wurden unter der Annahme multivariater Normalverteilung gerechnet. FIML berechnet fallweise eine *likelihood*-Funktion

mit den verfügbaren Daten. Es werden alle Daten genutzt, ohne die fehlenden zu ersetzen. Somit wird die Verteilungsfunktion verglichen mit MCAR und MAR besser genutzt. Eine unverzerrte Schätzung ist jedoch nur unter zwei Bedingungen möglich, nämlich MCAR und MAR. Liegen MCAR-Fehlerwerte vor, sind die beobachteten Variablen unabhängig von anderen beobachteten Variablen und auch von den Werten der Variablen mit den Fehlerwerten selbst, also rein zufällig. Liegen MAR Fehlerwerte vor, was einer Abschwächung des MCAR-Kriteriums gleichkommt, können die fehlenden Werte der Variablen zwar von einer anderen Variablen abhängen, jedoch nicht von der Variablen, in welcher die fehlenden Werte gerade auftauchen (Rubin, 1976). Für den listenweisen Fallausschluss muss MCAR vorliegen. Beantworten in der vorliegenden Studie beispielsweise nur jene Personen die weiteren Fragebögen, welche bei der ersten Befragung zufrieden waren, also MAR, würde das Ergebnis hinsichtlich dieser Personen verzerrt werden, da jene Personen, welche unzufrieden sind, aus der Analyse ausgeschieden werden. Beim paarweisen Fallausschluss werden ebenfalls MCAR-Daten benötigt, da die Werte der Kovarianzmatrix durch die paarweise Berechnung aus selbigem Grund wie beim listenweisen Fallausschluss verzerrt werden. Bei der *similar response pattern imputation*, bei welcher fehlende Werte durch vorhandene Werte ähnlicher Fälle ersetzt werden, werden stabile Schätzungen unter MCAR-Bedingungen gefunden. ML-Ansätze sollten sowohl für MCAR, als auch für MAR funktionieren.

Die Bevorzugung der Nutzung der gesamten Dateninformation führen Muthén et al. (1987) aus. Sie vergleichen anhand eines künstlichen Datensatzes drei unterschiedliche Lösungen in Fällen wo MCAR nicht angenommen werden kann, *full quasi-likelihood* (FQL), *listwise present (deletion) quasi-likelihood* (LQL) und *pairwise present approach* (PPA). Vergleiche werden anhand eines longitudinalen SEMs durchgeführt, wobei fehlende Werte aufgrund von Probandenausfall entstehen. FQL ordnet die Probanden in Gruppen mit gleichen Fehlerwertpatterns, wobei die Schätzungen mehrerer Gruppen simultan vorgenommen werden. Diese Lösung wird mit den Problemen der traditionellen Lösungen verglichen. PPA führt teilweise zu negativen Kovarianzmatrizen, LQL schätzt nur unverzerrt bei Vorliegen von MCAR. Gründe für fehlende Werte werden aus unterschiedlichen Sichtweisen getestet. Sie werden dargestellt durch sechs zum Zeitpunkt eins beobachtete Items (MAR), durch alle neun latente Responsevariablen (sechs beobachtete Variablen inklusive der drei später fehlenden), nur durch latente Variablen anstatt der beobachteten (beispielsweise durch die Selbstauswahl beim zweiten Test mitzumachen, was jedoch nicht durch die Punkte auf dem ersten Test oder durch kritische Fragen im ersten Test

beeinflusst ist) und durch Einbringung weiterer Variablen wie beispielsweise den sozioökonomischen Status (*marginal vs. full model*). In allen Fällen ist FQL zu bevorzugen, selbst wenn MCAR nicht gewährleistet werden kann, was meistens der Fall ist, selbst in Situationen, wo selbst MAR nicht mehr garantiert werden kann.

Alternativen zu LGC-Modellen stellen *exploratory longitudinal factor analysis*, *longitudinal SEM*, autoregressive Modelle und autoregressive *cross-lagged* Modelle oder auch Mehrebenenanalysen (Preacher et al., 2008) dar. Die Verknüpfung zu den letztgenannten Modellen, welche in dieser Arbeit bereits durchgeführt wurden, wird nochmals kurz angesprochen. *Hierarchical* oder *multi-level modeling* stellt eine mächtige Alternative zur Analyse latenter Wachstumskurven in SEM dar (Acock und Li, 1996). Mehrebenenanalysen, auch *hierarchical linear models*, *mixed models*, *variance component models* oder *random coefficient models* genannt, erlauben die Aufteilung der Varianz auf unterschiedliche genestete Levels. Beobachtungen sind innerhalb von Personen genestet. Ein Teil der Varianz wird durch die Beobachtungen einer Person und die Streuung über die Zeitpunkte bestimmt, *within-individuals* auf Level 1, der andere Teil durch die Varianz auf Individuenlevel, *between-individuals* auf Level 2. *Fixed coefficients* ignorieren die Variabilität auf Level 2, *random coefficients* schätzen die Variabilität zwischen den untersuchten Subjekten. Wenn die fehlenden Werte *missing at random* sind, können die vorhandenen Beobachtungen trotzdem verwendet werden. Dies ist ein großer Vorteil der MLM. Die Schätzung kann über FIML oder ML erfolgen. Die Software MLwiN, welche in der vorliegenden Arbeit verwendet wurde, benützt zur Bestimmung der ML Lösung *iterative generalized least squares* bzw. wenn dieser auf die Kleinst-Quadrat-Fehler angewandt wird, den *restricted iterative generalized least squares* (REML). Für ordinale Daten wie im vorliegenden Fall sind REML oder FIML nicht tauglich. Lösungen dafür wären *maximum marginal likelihood*, *penalized quasi likelihood*, *empirical Bayes* oder *Markov chain Monte Carlo* (Preacher et al., 2008). Werden lineare Modelle der LGM und der MLM gleich spezifiziert, liefert die Lösung gleiche Schätzungen für den *intercept*- und *slope*-Mittelwert, Varianzen, Kovarianzen und Fehlervarianzen. Unterschiede ergeben sich hinsichtlich der Spezifizierung des Messmodells in LGM. Beispielsweise kann für eine wiederholt gemessene Variable ein Fehlerterm spezifiziert werden, was vor allem für latente Variablen mit mehreren Indikatoren brauchbar ist. Dies ist in MLM sehr schwer umsetzbar, ebenso wie die Spezifizierung geschätzter Parameter als Prädiktoren. In LGM stellen diese Variablen dar. In MLM sind die Zeitpunkte bekannt, in LGM werden diese in den *unspecified growth*-Modellen frei geschätzt. Im Gegensatz zu LGM werden mehrere Hierarchien in MLM

leicht spezifiziert. Verbesserte Fehlerwertlösungen in LGM kommen an die Ergebnisse der MLM heran. Die Möglichkeit der Fehlerstrukturspezifizierungen in MLM nähert sich den Vorteilen der LGM an (Preacher et al., 2008).

4.3.6.3. *Latent growth curve model mit linearem slope - Kinodaten*

Hier ist das erste Ziel herauszufinden, ob sich die Bewertungen eines bestimmten Standortes, gleichgültig wann diese das erste, zweite oder dritte Mal in den ersten sechs Monaten stattfindet, über die Zeit hinweg verändern. Können positive Veränderungen aufgefunden werden, bewerten die Kinobesucher einen bestimmten Standort aufgrund von Servicequalitätsverbesserungen seitens der Kinobetreiber, oder aufgrund von milderer Bewertungen seitens der Besucher, etc., über die Befragungszeit hinweg besser als zu Beginn. Negative Veränderungen stehen für sinkende Bewertungen aufgrund von Veränderungen hinsichtlich schlechterer Serviceleistung eines Kinostandortes oder kritischerer Urteile der Zufriedenheitsaspekte, etc. Gleichbleibende Zufriedenheitsbewertungen signalisieren weder Veränderungen in der Beurteilung aufgrund von Änderungen den Standort betreffend, noch Veränderungen in der Beurteilung aufgrund der damit einhergehenden Konfrontation und Befassung mit der Befragungssituation im Zuge des Panels. Aber auch eine Aufhebung positiver und negativer Einwirkungen löst einen gleichbleibenden Zustand aus. In Tabelle 54 findet sich eine Auflistung der Anzahl an wiederholten Bewertungen.

Zuerst werden lediglich jene Personen in die Kalkulation miteinbezogen, welche in den ersten sechs Monaten mindestens dreimal ein Kino besucht haben. Diese drei Besuche müssen einen gleichbleibenden Standort betreffen. Zwischen den befragten Personen können die Standorte im Modell variieren. Der erste Besuch des speziellen Standortes, gleichgültig ob dieser in Welle 1, 2, 3 oder 4 war, zählt als Besuch zum Zeitpunkt eins. Der zweite Besuch des speziellen Standortes gleichgültig ob dieser in Welle 2, 3, 4 oder 5 war, zählt als Besuch zum Zeitpunkt zwei. Der dritte Besuch des speziellen Standortes gleichgültig ob dieser in Welle 3, 4, 5 oder 6 war, zählt als Besuch zum Zeitpunkt drei. Die maximal dreimal wiederholten Bewertungen eines bestimmten Kinos werden zurück verschoben und ersetzt dort auftauchende Bewertungen eines anderen Kinostandortes bzw. fehlende Bewertungen. Es wird von einem gleichbleibenden Level, *intercept*, und einem linearen Wachstumsparameter, *slope*, welcher gleichbleibende Veränderungen über die drei Perioden ausgehend vom *intercept* misst, in Diagramm 117 ausgegangen.

Anzahl an unterschiedlichen Kinos	Einzelbesuche						Zweifachbesuche						...					
	1x	2x	3x	4x	5x	6x	1x	2x	3x	4x	5x	6x	1x	2x	3x	4x	5x	6x
1	1918						1918											
2	1264						2528											
3	576						1728											
4	211						844											
5	56						280											
6	16						96											
1			1623					1623										
2			239					478										
3			2					6										
1				829					829									
2				14					28									
1					393					393								
1						150					150							
1							47					47						
							10948						10948	mind 1				
								3554					3554	mind 2				
									1447				1447	mind 3				
										590			590	mind 4				
											197		197	mind 5				
												47	47	mind 6				
							7394	2107	857	393	150	47						
							nur 1	nur 2	nur 3	nur 4	nur 5	nur 6						

Bewertungen für eine bestimmte Anzahl an Wiederholungen
 Tabelle 54: Übersicht über die Anzahl an Wiederholungen für bestimmte Kinos

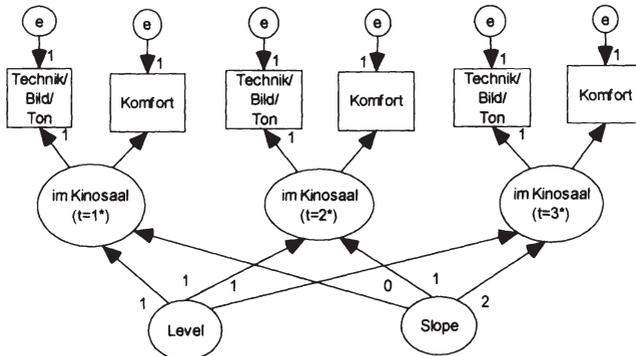


Diagramm 117: Lineares LGC-Modell

Die Veränderungen der im Kapitel Faktorenbestimmung ermittelten drei latenten Dimensionen werden anhand von 1.446 Beobachtungen ohne Ausschluss von Personen mit fehlenden Werten untersucht. In Klammern findet sich der *F*-Wert. Die latente Dimension ‚vor dem Kinobesuch‘ weist eine knapp nicht signifikante negative Veränderung von -0,008 (-1,868) auf, ‚im Kino‘ eine positiv signifikante von 0,033 (3,775) und ‚im Kinosaal‘ eine nicht signifikante von -0,001 (-0,103). Der *slope*-Wert des Faktors ‚im Kino‘, welcher Auswirkungen auf die Zufriedenheitsindikatoren Personal, Preis/Leistung, Buffet und Atmosphäre hat bedeutet, dass jeder Bewertung im Durchschnitt pro Zeitsprung ein Zufriedenheitswert auf einer 6-stufigen Likert-Skala von 0,033 hinzugezählt werden muss. Die befragten Personen bewerten diese Zufriedenheitskriterien über die Zeit besser. Der Faktor ‚vor dem Kinobesuch‘, welcher sich auf Standort, Filmangebot, Informationsangebot, Reservierungsmöglichkeiten und Erscheinungsbild auswirkt, weist eine knapp signifikant kritischere Bewertungsauswirkung auf und die Sensibilität bezüglich dieser Zufriedenheitsindikatoren steigt somit. Die latente Dimension ‚im Kinosaal‘ bewirkt keine Veränderungen der Indikatoren Technik/Bild/Ton und Komfort. Werden die Personen mit Fehlerwerten ausgeschlossen, finden sich 1.087 Beobachtungen für die *slope*-Messung des Faktors ‚vor dem Kinobesuch‘ von -0,007 (-1,525). 890 Beobachtungen zeigen eine signifikant positive Veränderung des Faktors ‚im Kino‘ von 0,026 (2,545) und 1.426 Beobachtungen und keine signifikante von 0,000 (0,044) für ‚im Kinosaal‘. Die beobachteten Items können auch als ordinal-skalierte (kategoriale) Variablen behandelt werden. Ohne Ersatz der Fehlerwerte ist lediglich ein Wachstumsfaktor bei 1.426 Beobachtungen für ‚im Kinosaal‘ mit den Werten 0.042 (0.011) schätzbar. Werden die fehlenden Werte ersetzt, zeigt sich bei 1.446 Probanden ein nicht signifikanter *slope* von 0.036

(0.018). Die weiteren beiden Dimensionen ‚vor dem Kino‘ und ‚im Kino‘ können nicht geschätzt werden.

Für die Rechtfertigung des Modells, muss gewährleistet werden, dass die Ladungen der Faktoren auf die manifesten Items sich über die Zeit nicht ändern. Eine Strukturänderung der einzelnen Ladungen zwischen zwei Zeitpunkten bedeutet, dass der jeweilige latente Faktor zum ersten Zeitpunkt eine andere Bedeutung hat als zum nächsten Zeitpunkt. In Tabelle 55 werden die über drei Wiederholungen gleichgesetzten Ladungen der drei Faktoren auf die Items dargestellt. Alle latenten Faktoren laden signifikant auf die zugehörigen Items. Jeweils die erste Ladung des latenten Konstruktes auf die manifesten Variablen wird auf eins fixiert.

Item	im Kinosaal	im Kino	vor dem Kinobesuch
Technik/Bild/Ton	1,000 (0,000)		
Komfort	1,616 (24,477)		
Servicepersonal		1,000 (0,000)	
Preis/Leistung		1,068 (23,783)	
Buffet		1,400 (28,469)	
Atmosphäre		0,907 (26,449)	
Standort/Erreichbarkeit			1,000 (0,000)
Filmangebot			1,837 (19,949)
Informationsangebot			2,030 (20,571)
Reservierungsmöglichkeiten			1,416 (16,251)
Erscheinungsbild/Image			1,691 (19,761)

Tabelle 55: Ladungsstruktur

Um verstärkt das allgemeine Zufriedenheitsempfinden, als auch Panelbefragungsveränderungen aufgrund möglicher einhergehender kritischerer Betrachtung der Fragen mit einzubeziehen, werden alle Bewertungen verwendet, gleichgültig welcher Standort bewertet wurde. Die Ergebnisse finden sich in Tabelle 56. Nicht schätzbare Werte werden mit ‚-‘ ausgewiesen.

Werden nur die Veränderungen eines bestimmten Standortes gemessen, werden nur jene Fälle inkludiert von welchen es mindestens drei Wiederholungen einen bestimmten Standort betreffend gibt. Dieser Standort ist auch über die Personen hinweg gleich. Dafür wird ermittelt, welche der Standorte die größte Bewertungshäufigkeit aufweisen. Dies sind #3 mit 210 Beobachtungen, #15 mit 147 Beobachtungen, #16 mit 227 Beobachtungen, #18 mit 294 Beobachtungen

und #48 mit 229 Beobachtungen. Die Ergebnisse unter Betrachtung der ersten drei Wiederholungen und Nicht-Ausschluss der Personen mit fehlenden Werten finden sich in Tabelle 57. In eckigen Klammern findet sich die Anzahl an Beobachtungen.

Wiederholungen	Fehlende Werte	Faktor	Beobachtungen	Slope	Signifikanzwert
3	Berücksichtigt	Vor dem Kinobesuch	5.145	-0,019	-3,618
		Im Kino		0,027	3,346
		Im Kinosaal		-0,002	-0,221
6	Nicht berücksichtigt	Vor dem Kinobesuch	961	-0,009	-1,057
		Im Kino	845	0,039	2,725
		Im Kinosaal	1.529	0,000	0,041
3	Berücksichtigt	Vor dem Kinobesuch	5.548	-0,013	-6,390
		Im Kino		0,019	6,282
		Im Kinosaal		-0,005	-1,809
6	Nicht berücksichtigt	Vor dem Kinobesuch	286	-0,002	-0,376
		Im Kino	248	-	-
		Im Kinosaal	555	-0,004	-0,551

Tabelle 56: Linearer slope – allgemein

	Kino #3 [210]	Kino #15 [147]	Kino #16 [227]	Kino #18 [294]	Kino #48 [229]
Vor dem Kinobesuch	-	-	-0,025 (-1,425)	-	-
Im Kino	-	0,072 (2,550)	0,024 (1,032)	-	-
Im Kinosaal	-0,015 (-0,718)	0,033 (1,197)	-	-	-

Tabelle 57: Linearer slope - standortspezifisch

Lediglich bei Standort #15 findet sich ein positiv signifikanter Wachstumsfaktor für ‚im Kino‘. Dies bedeutet, dass bei Standort #15 die Zufriedenheitsbewertungen der latenten Dimension über die Zeit hinweg positiver ausfallen. In den anderen Fällen ist das Modell entweder nicht schätzbar oder es finden sich keine signifikanten Wachstumsveränderungen. Die Aufspaltung des Datensatzes nach Standorten bringt somit keine wirkliche Verbesserung. Dies kann auch mit der geringen Anzahl an Beobachtungen zusammenhängen.

4.3.6.4. Latent growth curve model mit quadratischem slope - Kinodaten

Im folgenden Wachstumsmodell wird ein gleichbleibender Level, *intercept*, angenommen und ein Wachstumsparameter, *slope*, misst eine quadratische Veränderung ausgehend von diesem Level über drei Perioden hinweg. Die Ladung der ersten Bewertung wird als Ausgangspunkt verwendet. Die zweite Bewertung erhält eine *slope*-Ladung von eins und definiert somit die Größe der Veränderung über ein Zeitintervall. Die Ladung der dritten Bewertung enthält den gleichen *slope*, aber der im Modell theoretisch implementierte Zeitabstand wird auf $2^2=4$ festgelegt. Somit wird wie in Diagramm 118 gezeigt, weiterhin eine lineare Steigung geschätzt, jedoch der Zeitabstand auf ein vierfaches der durch den Abstand zwischen der ersten und der zweiten definierten Zeitspanne fiktiv erhöht. Wird das Ergebnis im Anschluss zusammengestaucht, befindet sich zwischen Bewertung zwei und drei die vierfache absolute Steigung des zwischen der ersten und zweiten Bewertung ermittelten *slopes*. Wird ein quadratischer *slope* über vier Wellen geschätzt, befindet sich zwischen der dritten und vierten Bewertung ein modellierter Zeitabstand von $3^2=9$ Zeitintervallen und somit der 9-fache *slope*-Wert. Es wird davon ausgegangen, dass es in den ersten drei Bewertungen die stärksten Veränderungen gibt, da hier verstärkt die kritischere Betrachtung aufgrund der wiederholten Befragung zu Tage tritt. Weder mit noch ohne Ersatz fehlender Werte kann ein signifikant quadratischer *slope* nachgewiesen werden. Eine derartig starke Bewertungsveränderung kann somit nicht identifiziert werden.

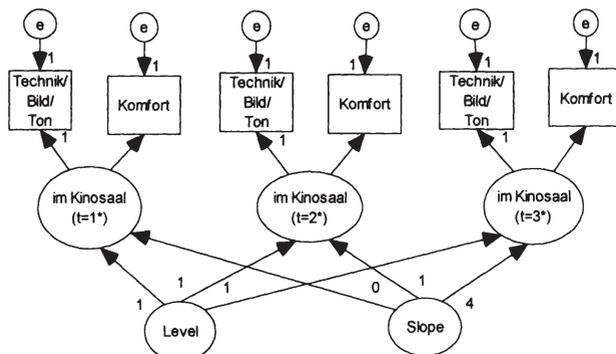


Diagramm 118: Quadratisches LGC-Modell

4.3.6.5. Latent growth curve model mit frei geschätzten slope-Parametern - Kinodaten

Um den aus der Datenstruktur vorgeschlagenen *slope* zu identifizieren, wird der Zeitabstand, welcher im zuvor modellierten quadratischen Modell zwischen

Bewertung zwei und drei auf vier Zeitintervalle definiert wurde, wie in Diagramm 119 mit ‚*‘ gekennzeichnet, frei geschätzt. Fällt der frei geschätzte Parameter unter zwei, steht dies für eine abgeschwächte Veränderung, im Gegensatz zur Veränderung zwischen Bewertung eins und zwei im linearen Modell. Erhält der Term einen Wert über zwei, wird zwischen zweiter und dritter Bewertung eine stärkere Veränderung als linear fortgeschrieben.

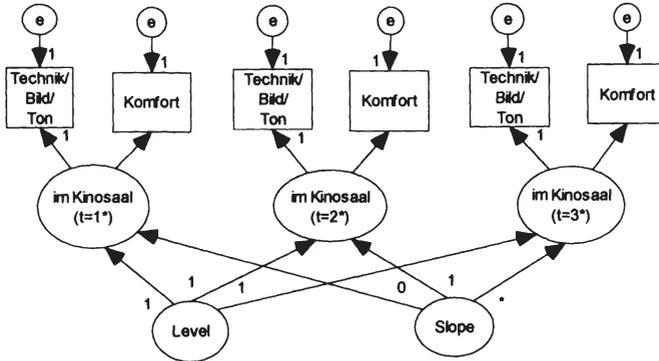


Diagramm 119: Frei geschätztes LGC-Modell

In Tabelle 58 finden sich die Ergebnisse des Modells ohne Berücksichtigung der Probanden mit fehlenden Werten. Für ‚im Kinosaal‘ mit 1.426 Bewertungen kann kein freier Term geschätzt werden. Für ‚im Kino‘ wird davon ausgegangen, dass der eigentliche Wert zwar nicht an eine quadratische Veränderung herankommt, sehr wohl jedoch stärker als der lineare Zusammenhang ist. ‚Vor dem Kinobesuch‘ fällt unter den linearen Trend. Jedoch ist keiner der *slopes* signifikant. Es findet sich eine kleine Abweichung gegenüber dem linearen Modell. Im Vergleich zum linearen Modell, in welchem für die Dimension ‚vor dem Kinobesuch‘ ein *slope* von -0.008 (-1.868) und für ‚im Kino‘ 0.004 (0.829) geschätzt wurde, bringt die freie Schätzung keine Veränderungen in der Signifikanz der *slopes*.

Faktor	Beobachtungen	Zeitabstand	F-Wert	Slope	F-Wert
Vor dem Kinobesuch	1.087	1,169	1,536	-0,013	-1,339
Im Kino	890	2,553	1,186	0,020	0,991

Tabelle 58: Frei geschätzter slope

Ob das lineare Modell gegenüber dem frei geschätzten verworfen werden muss, wird anhand der Modellfitwerte für ‚vor dem Kinobesuch‘ und ‚im Kino‘ in

Tabelle 59 überprüft. Es finden sich keine signifikanten Unterschiede und das schmalere lineare Modell wird bevorzugt.

Beurteilung	Vor dem Kinobesuch		Im Kino	
	Linear	Frei	Linear	Frei
Schätzung				
Chi-Square Test of Model Fit	3.047,172	3.046,720	1.249,576	1.249,465
Chi-Square Differenz	0,452		0,111	
Freiheitsgrade	104	103	64	63
Freiheitsgrade Differenz	1		1	
p-Wert	0,0000	0,0000	0,0000	0,0000
Wahrscheinlichkeit/Wert/Bewertung	<0,7/1,07/nicht signifikant		<0,7/1,07/nicht signifikant	

Tabelle 59: Lineares vs. frei geschätztes LGC-Modell

4.3.6.6. *Piecewise latent growth curve model* - Kinodaten

Um zu testen, ob in den ersten Bewertungen eine stärkere Veränderung aufgrund der Befragung stattfindet, welche sich in den Folgewellen abschwächt, werden zwei unterschiedliche *slopes* geschätzt. Erster *slope* misst die Veränderung zwischen erster und zweiter und zweiter und dritter Bewertung. Zweiter *slope* hat als Ausgangspunkt den *intercept*-Wert der dritten Bewertung und misst dazu die Veränderung zwischen Bewertung drei und vier, und vier und fünf. Es werden alle Daten der zwölf Befragungswellen verwendet und die ersten fünf Bewertungen des meistbesuchten Kinostandortes zur Betrachtung herangezogen. Die Datenaufbereitung erfolgt gleich wie im obigen Fall, jedoch unter Verwendung aller Befragungswellen. Zuerst wird das Modell mit linearem *slope* und konstanten Zeitabständen für einen anschließenden Vergleich in Tabelle 60 geschätzt.

Wiederholungen	5 (Probandenausschluss mit fehlenden Werten)			5		
	Im Kinosaal	Im Kino	Vor dem Kinobesuch	Im Kinosaal	Im Kino	Vor dem Kinobesuch
Probanden	1.220	695	892	1.247		
Slope	-0,005	0,010	-0,005	-0,006	0,011	-
F-Wert	-1,457	1,844	-2,151	-1,597	2,527	-

Tabelle 60: Lineares LGC-Modell – 5 Wellen

Für ‚vor dem Kinobesuch‘ wird ein leicht signifikanter negativer *slope* und somit sinkende Zufriedenheit verzeichnet, für ‚im Kino‘ wird dieser teilweise gezeigt. Das Modell wird in Diagramm 120 dargestellt.

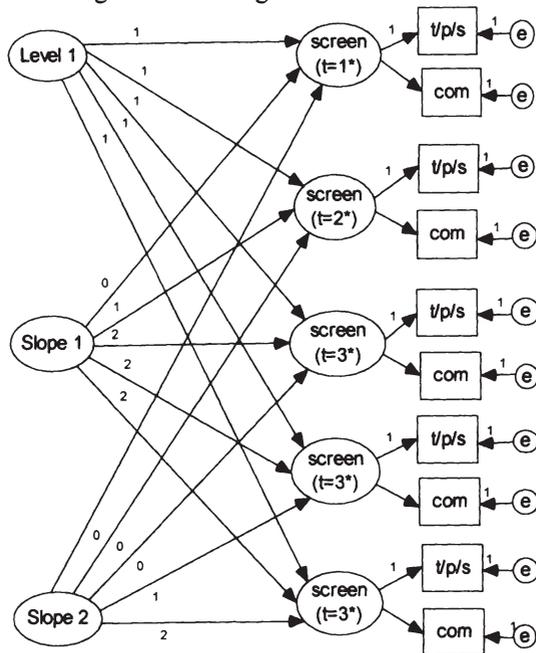


Diagramm 120: Piecewise LGC-Modell

Nun wird geprüft, ob die Veränderung konstant über die Zeit ist, oder variiert. Beide *slopes* sind ausschlaggebend. *Slope 1* zeigt die lineare Veränderung zwischen Bewertung eins und zwei, und zwei und drei, *slope 2* die zwischen drei und vier, und vier und fünf. In Tabelle 61 zeigt sich keine sinnvolle Schätzung für einen zweiten linearen *slope* zwischen Bewertung drei und vier, und vier und fünf. Es findet sich jedoch ein signifikanter *slope* für ‚im Kino‘, falls Probanden mit fehlenden Werten berücksichtigt werden und für ‚vor dem Kinobesuch‘ (in beiden Fällen) zwischen erster und zweiter, und zweiter und dritter Bewertung.

Es zeigt sich weiters für den latenten Servicefaktor ‚im Kino‘ eine signifikant negative Korrelation zwischen beiden *slopes*, für das Modell in welchem Probanden mit fehlenden Werten nicht ausgeschlossen wurden. Je höher also *slope 1* ist, desto niedriger *slope 2*. Ein Proband dessen Antworten zwischen Bewertung eins und drei einer starken Verbesserung unterliegen, dessen Antworten beinhalten zwischen Bewertung drei und fünf eine abgeschwächte

Verbesserung. Die gegenteilige Interpretation des Korrelationskoeffizienten führt bei einem niedrigen *slope 1* zu einem höheren *slope 2*. Somit wird ein Proband, dessen Antworten zwischen Bewertung eins und drei einer schwachen Verbesserung unterliegen, mehr Spielraum zur Verbesserung in den Folgewellen haben und zwischen Bewertung drei und fünf eine verstärkte Verbesserung angeben.

Wiederholungen	5 (Probandenausschluss mit fehlenden Werten)		
Faktor	Im Kinosaal	Im Kino	Vor dem Kinobesuch
Probanden	1.220	695	892
Slope 1 (F-Wert)	-0,004 (-0,703)	-	-0,009 (-2,066)
Slope 2 (F-Wert)	0,000 (0,000)	-	0,000 (0,000)
Korrelation Slope 1 und 2 (F-Wert)	-0,002 (-0,778)		0,000 (0,460)
Wiederholungen	5		
Probanden	1.247		
Slope 1 (F-Wert)	-0,005 (-0,856)	0,029 (3,760)	-0,008 (-2,203)
Slope 2 (F-Wert)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
Korrelation Slope 1 und 2 (F-Wert)	-0,002 (-0,876)	-0,016 (-3,803)	0,000 (0,546)

Tabelle 61: Piecewise LGC-Modell

4.3.6.7. *Parallel latent growth curve model* mit Effekten - Kinodaten

Dieser Umstand wird anhand einer weiteren Konstellation in Diagramm 121 verdeutlicht. Wiederum wird der gesamte Datensatz aller zwölf Wellen verwendet, jedoch ebenfalls die Bewertungen standortspezifisch erfasst. Es werden zwei *slopes* und zwei *intercepts* geschätzt. *Slope 1* und *intercept 1* werden für die Bewertungen eins auf zwei, und zwei auf drei geschätzt, *Slope 2* und *intercept 2* für die Bewertungen drei auf vier, und vier auf fünf. Es werden sechs latente Konstrukte für die zwei Serviceleistungsisems errechnet und vier Wachstumsfaktoren zur Beeinflussung dieser. Die Schätzung findet keine sinnvollen Ergebnisse. Aus der Veranschaulichung könnten Schlüsse zum Zusammenhang zwischen *intercept* und *slope* gezogen werden. Wird ein Proband, welcher zu Beginn gemäßigte Antworten gibt, einer schwächeren Veränderung unterzogen, im Vergleich zu einem Probanden, welcher zu Beginn extremere Antworten gibt? Fällt die Bewertung von Probanden über die Zeit hinweg in den negativ kritischen Bereich, oder verändern sich positiv antwortende Probanden ins positive Extreme und negativ antwortende ins negativ Extreme?

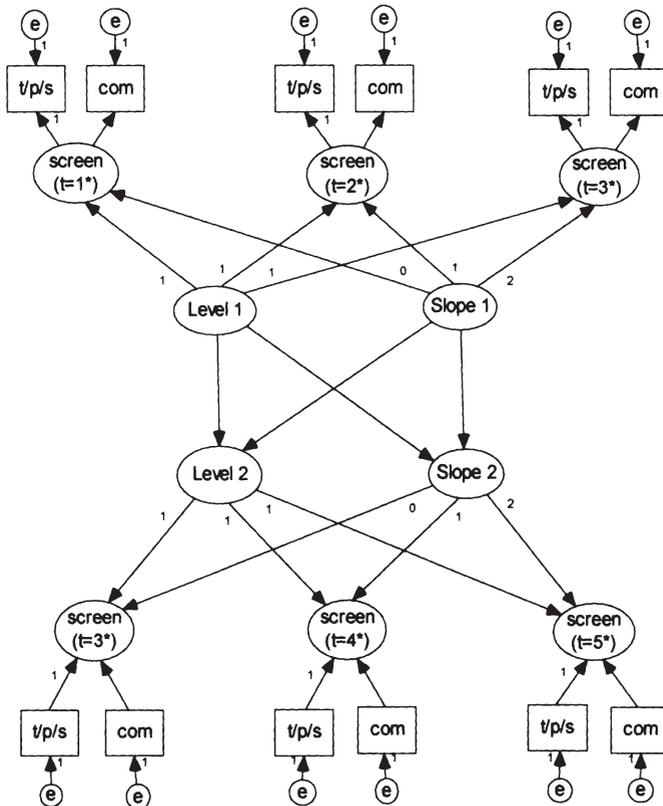


Diagramm 121: Erweiterung zu getrennten Wachstumsverläufen

4.3.6.8. Kinostandortspezifische Veränderungen und Fehlerwertverzerrungen - Kinodaten

Nun werden jene Personen in die Schätzung miteinbezogen, welche in den ersten zwölf Monaten mindestens dreimal ein Kino besucht haben. Diese drei Besuche betreffen einen gleichbleibenden Standort. Zwischen den befragten Personen variieren die Standorte. Zusammengefasst werden die maximal dreimal wiederholten Bewertungen eines bestimmten Kinos zurück verschoben und anstatt der Bewertungen eines anderen Kinostandortes oder fehlenden Werten gesetzt. Die Form der Aufbereitung wird auch für Kinostandorte bei einer wiederholten Bewertung von 4, 5, 6 und 7 vorgenommen und in Tabelle 62 dargestellt. Ziel ist die Aufdeckung von Bewertungsveränderungen eines bestimmten Standortes über die Zeit, gleichgültig wann diese das erste, zweite, dritte, etc. mal in den zwölf Monaten stattfindet.

Im folgenden LGC-Modell misst ein Wachstumsparameter, *slope*, eine lineare gleichbleibende Veränderung über drei bis sieben Perioden hinweg. Tabelle 63 zeigt eine Ergebnisübersicht. Gleichgültig ob die Fehlerwerte ersetzt werden oder nicht, ergeben sich beinahe die gleichen Ergebnisse. ‚Vor dem Kinobesuch‘ ist über vier und fünf Wellen signifikant negativ und wird somit im Zeitablauf schlechter bewertet. Der *slope* fällt zwischen drei und fünf aufeinanderfolgenden Beobachtungen ab und verliert an Signifikanz. ‚Im Kino‘ weist bei vier und fünf Wiederholungsmessungen einen positiv signifikanten *slope* auf. Probanden bewerten diesen latenten Faktor über die Zeit hinweg besser. Diese Verbesserung kann nicht über sechs bzw. sieben Wiederholungen gezeigt werden. ‚Im Kinosaal‘ ist über vier Wellen signifikant. Der Signifikanzwert sinkt über die Verlängerung der Wachstumsmessdauer und ist über sieben Wiederholungen nicht mehr schätzbar.

Fehlerwertbehandlung - Kein Probandenausschluss						
Wiederholung		3	4	5	6	7
Im Kinosaal	Probanden	-	1,891	1,247	771	-
	Slope	-	-0,010	-0,006	-0,004	-
	F-Wert	-	-2,826	-1,597	-1,291	-
Im Kino	Probanden	2,894	1,891	1,247	771	-
	Slope	0,001	0,022	0,011	0,005	-
	F-Wert	0,009	3,655	2,527	1,268	-
Vor dem Kinobesuch	Probanden	-	1,891	1,247	-	-
	Slope	-	-0,029	-0,011	-	-
	F-Wert	-	-4,910	-2,180	-	-
Fehlerwertbehandlung - Probandenausschluss						
Wiederholung		3	4	5	6	7
Im Kinosaal	Probanden	2,834	1,847	1,220	753	460
	Slope	-	-0,010	-0,005	-0,004	-
	F-Wert	-	-2,033	-1,457	-1,059	-
Im Kino	Probanden	1,649	1,052	695	455	276
	Slope	-0,001	0,021	0,010	0,004	-
	F-Wert	-0,050	2,816	1,844	0,829	-
Vor dem Kinobesuch	Probanden	1,994	1,338	892	563	349
	Slope	-0,030	-0,011	-0,005	-	-
	F-Wert	-4,241	-2,616	-2,151	-	-

Tabelle 63: Wiederholungsmessungen

Die Zufriedenheitsvariablen können auch als ordinal skalierte (kategoriale) Variable behandelt werden. Die Ergebnisse für drei Wiederholungen unter Ausschluss von Personen mit fehlenden Werten liefern lediglich einen Wachstumsfaktor für ‚im Kinosaal‘ -0.389 (-0.488) bei 2.834 Probanden. Werden Probanden mit fehlenden Werten berücksichtigt, ergibt sich ein nicht signifikanter *slope* von 0.099 (0.095) bei 2.894 Probanden. Die Modelle der weiteren Dimensionen sind nicht schätzbar. Deshalb werden die Items in den folgenden Modellen wieder als intervallskaliert behandelt, um die nicht mögliche Schätzung der *thresholds* zu vermeiden.

Zur Überprüfung ob es standortspezifisch unterschiedliche Veränderungen gibt, werden Wachstumsparameter für die einzelnen Standorte berechnet. Aufgrund der geringen Fallzahlen für sechs bzw. sieben Wiederholungen, wird ein Vergleich unterschiedlicher Kinostandorte für Veränderungen über fünfmalige Bewertungen vier unterschiedlicher Standorte betrachtet. Tabelle 64 gibt Aufschluss über kinostandortspezifische Veränderungen. Es wird lediglich ein einziger signifikanter Wachstumsfaktor auf kinostandortspezifischer Ebene aufgefunden. Im Kino #18 tendieren die Probanden dazu, sinkende Bewertungen für ‚im Kinosaal‘ in den ersten fünf Bewertungen dieses Kinos abzugeben. Der Wachstumsfaktor kann jedoch nicht für sechs Wiederholungsmessungen bei 185 (187) Probanden bzw. sieben bei 117 (119) Probanden nachgewiesen werden. In Klammern stehen die Probandenzahlen wenn Personen mit fehlenden Werten nicht ausgeschlossen werden. Es können durch standortspezifisch genauere Messungen keine unterschiedlichen *slopes* zwischen verschiedenen Standorten über einen längeren Bewertungszeitraum nachgewiesen werden.

Die Ergebnisse mit frei geschätzten Zeitkoeffizienten finden sich in Tabelle 65 für vier Wiederholungsbewertungen. Für ‚im Kinosaal‘ wird kein signifikanter *slope* gefunden. ‚Im Kino‘ zeigt einen fast signifikanten *slope* über vier Messwiederholungen, welcher eine sinkende Tendenz aufweist. Wird ein linearer Verlauf geschätzt, liegen die Ladungswerte bei 0, 1, 2 und 3 für die vier Zeitpunkte. Hier liegen diese bei 0, 1, 2,152 und 1,713. Dies bedeutet, dass die Veränderung zwischen zweiter und dritter Wiederholung positiv stärker ist, verglichen zur der zwischen eins und zwei. Sie sinkt jedoch wieder zwischen Bewertung drei und vier. ‚Vor dem Kinobesuch‘ zeigt zwischen Bewertung zwei und drei ein stark negativen Trend, welcher später ebenfalls abfällt.

Fehlerwertbehandlung - Kein Probandenausschluss					
Kino Nr.		3	16	18	48
Im Kinosaal	Probanden	-	193	287	-
	Slope	-	-0,003	-0,013	-
	F-Wert	-	-0,374	-2,727	-
Im Kino	Probanden	158	193	-	-
	Slope	0,025	0,010	-	-
	F-Wert	1,700	0,870	-	-
Vor dem Kinobesuch	Probanden	-	-	-	-
	Slope	-	-	-	-
	F-Wert	-	-	-	-
Fehlerwertbehandlung - Probandenausschluss					
Kino Nr.		3	16	18	48
Im Kinosaal	Probanden	157	192	282	202
	Slope	-	-0,002	-0,013	-
	F-Wert	-	-0,25	-2,763	-
Im Kino	Probanden	77	107	182	115
	Slope	-	-	-	0,003
	F-Wert	-	-	-	0,237
Vor dem Kinobesuch	Probanden	112	137	223	149
	Slope	-	0,004	-	-
	F-Wert	-	0,466	-	-

Tabelle 64: Kinosppezifisch – fünf Wellen

Faktor	Ausschluss bei fehlenden Werten		
	Im Kinosaal	Im Kino	vor dem Kinobesuch
Probanden	-	-	1.338
Slope	-	-	-0,014 (-1,684)
Ladungen/Zeit-Koeffizienten	-	-	1,045 (2,436) 3,733 (1,941)
Faktor	Kein Ausschluss bei fehlenden Werten		
Probanden		1.891	1.891
Slope	-	0,032 (2,191)	-0,010 (-1,812)
Ladungen/Zeitkoeffizienten	-	2,152 (2,660) 1,713 (3,374)	1,396 (3,473) 4,612 (1,937)

Tabelle 65: Freie Schätzung - standortspezifisch

Der zuletzt erwähnte *slope* für ‚vor dem Kinobesuch‘ mit seinen zugehörigen Ladungen wird für fünf Wiederholungsmessungen geschätzt und in Tabelle 66 umgerechnet. Der freie *slope*-Parameter ist über sechs und sieben Bewertungszeitpunkte nicht mehr schätzbar.

Ausschluss bei Fehlerwerten (892 Probanden)				
	Wert	F-Wert	Berechnung	Ergebnis
Slope	-0,010	-1,945		
Ladung 1	0,000	0,000	$-0,01 \times 0,000$	0,0000
Ladung 2	1,000	0,000	$-0,01 \times 1,000$	-0,0100
Ladung 3	1,709	4,123	$-0,01 \times 1,709$	-0,01709
Ladung 4	2,251	3,826	$-0,01 \times 2,251$	-0,02251
Ladung 5	2,398	3,752	$-0,01 \times 2,398$	-0,02398
Kein Ausschluss bei Fehlerwerten (1.247 Probanden)				
	Wert	F-Wert	Berechnung	Ergebnis
Slope	-0,007	-1,919		
Ladung 1	0,000	0,000	$-0,007 \times 0,000$	0,000000
Ladung 2	1,000	0,000	$-0,007 \times 1,000$	-0,007000
Ladung 3	2,292	3,490	$-0,007 \times 2,292$	-0,016044
Ladung 4	2,836	3,192	$-0,007 \times 2,836$	-0,019852
Ladung 5	3,260	3,095	$-0,007 \times 3,260$	-0,022820

Tabelle 66: Freie Schätzung - Berechnung

Zeigt der freie Verlauf eine lineare Schätzung, müssen alle Differenzwerte zwischen den vier Zeitintervallen die Ladungen für den linearen Verlauf aufweisen. Dies würde den Werten 0, 1, 2, 3 und 4 entsprechen und die Differenz zwischen den Werten muss in jedem Zeitintervall gleich eins sein. Die Differenzwerte der frei geschätzten Parameter zeigen in Tabelle 67, dass dies nicht der Fall ist. Die Werte weichen eindeutig vom linearen Trend mit Differenzwerten von eins ab.

Nachfolgend wird die Veränderung beider Modelle, mit und ohne Ausschluss bei Auftreten von Fehlerwerten, in Diagramm 122 visualisiert. Die Verlaufskurve zeigt einen sinkenden nicht-linearen Verlauf. Zu Beginn wird ‚vor dem Kinobesuch‘ einer starken negativ kritischeren Bewertung unterzogen. In den späteren Zeitpunkten flacht dieser Trend ab.

	Hypothetisch linear			
	Berechnung	Differenz		
Bewertung 2 - 1	1,000-0,000	1,000		
Bewertung 3 - 2	2,000-1,000	1,000		
Bewertung 4 - 3	3,000-2,000	1,000		
Bewertung 5 - 4	4,000-3,000	1,000		
	Frei ohne Fehlerwertausschluss		Frei mit Fehlerwertausschluss	
	Berechnung	Differenz	Berechnung	Differenz
Bewertung 2 - 1	1,000-0,000	1,000	1,000-0,000	1,000
Bewertung 3 - 2	2,292-1,000	1,292	2,292-1,000	1,292
Bewertung 4 - 3	2,836-2,292	0,544	2,836-2,292	0,544
Bewertung 5 - 4	3,260-2,836	0,424	3,260-2,836	0,424

Tabelle 67: Differenzberechnung

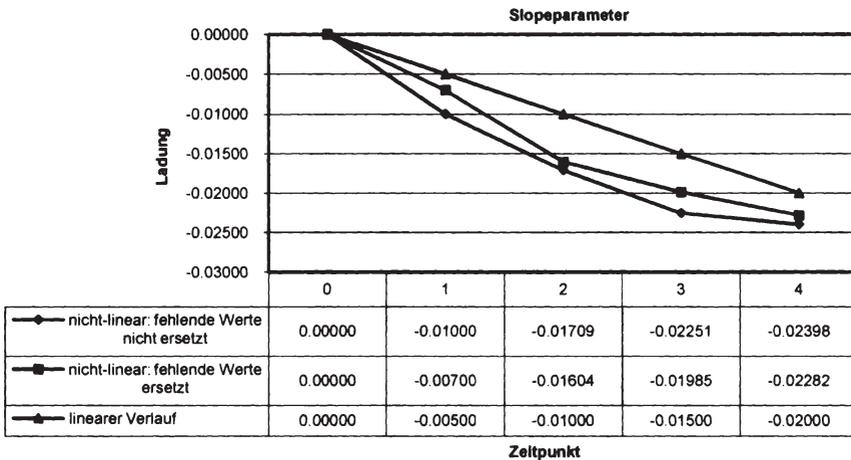


Diagramm 122: Linearer vs. nicht-linearer Verlauf

4.3.6.9. Lineare vs. nicht-lineare Wachstumsverläufe - Kinodaten

Nun werden die zwei Varianten verglichen, um zu bestimmen, ob der lineare Verlauf oder der nicht-lineare Verlauf über fünf Bewertungswiederholungen besser geeignet ist. Die Schätzungen für den nicht-linearen Verlauf mit Fallausschluss bei fehlenden Werten wurden bereits genannt. Der lineare Verlauf bei gleicher Probandenzahl weist einen signifikant gleichbleibenden *slope* über fünf Beobachtungen von -0.005 (-2.151) auf, wie auch im vorangegangenen Diagramm veranschaulicht. Mit einem χ^2 -Differenztest wird die Eignung der beiden Varianten verglichen. Der nicht-lineare Verlauf mit freier Schätzung

zeigt einen Chi-Square Test of Model Fit-Wert von 6439,924 bei 304 Freiheitsgraden. Der lineare Verlauf zeigt einen Chi-Square Test of Model Fit-Wert von 6445,045 bei 307 Freiheitsgraden. Die Differenz beträgt 5,121 bei 3 Freiheitsgraden und der Wert liegt laut χ^2 -Verteilung für 3 Freiheitsgrade zwischen einer Wahrscheinlichkeit von 0,8 bei einem Wert von 4,64 und 0,85 bei einem Wert von 5,32. Die Nullhypothese, dass der nicht-lineare Verlauf gleich gut ist wie der lineare kann nicht verworfen werden. Lediglich bei einer Irrtumswahrscheinlichkeit von etwas weniger als 20 bzw. 15 % wird die nicht-lineare Variante bevorzugt.

Um den Unterschied zwischen der linearen und nicht-linearen Lösung anschaulicher darzustellen, werden zuerst für beide Modelle mit 892 Probanden und Ausschluss bei Vorliegen von fehlenden Werten, in Tabelle 68 die 95%-igen Konfidenzintervalle berechnet (Mittelwert +/- 1,96*Standardabweichung).

	Linear	Berechnung	Unteres Konfidenzintervall (5%)	Oberes Konfidenzintervall (5%)
Linear	Bewertung 1	0,000	-	-
	Bewertung 2	-0,005 +/- 1,96*0,002	-0,00892	-0,00108
	Bewertung 3	-0,010 +/- 1,96*0,002	-0,01392	-0,00608
	Bewertung 4	-0,015 +/- 1,96*0,002	-0,01892	-0,01108
	Bewertung 5	-0,020 +/- 1,96*0,002	-0,02392	-0,01608
Nicht-linear	Bewertung 1	0,00000	-	-
	Bewertung 2	-0,01000 +/- 1,96*0,005	-0,01980	-0,00020
	Bewertung 3	-0,01709 +/- 1,96*0,005	-0,02689	-0,00729
	Bewertung 4	-0,02251 +/- 1,96*0,005	-0,03231	-0,01271
	Bewertung 5	-0,02398 +/- 1,96*0,005	-0,03378	-0,01418

Tabelle 68: Konfidenzintervallberechnung – linear und nicht-linear

Die geschätzten Parameterwerte und zugehörigen Konfidenzintervalle werden in Diagramm 123 dargestellt. An jenen Stellen, an welchen sich die Konfidenzintervalle der linearen und nicht-linearen Lösung überschneiden, gibt es visuell keine merkbaren Unterschiede. An jenen Stellen, an welchen sich diese nicht überschneiden, liegt ein visueller Unterschied vor. Aufgrund der hohen Abweichungen und folglich breiten Konfidenzintervallen, vor allem des nicht-linearen Verlaufes, ist kein visueller Unterschied nachweisbar.

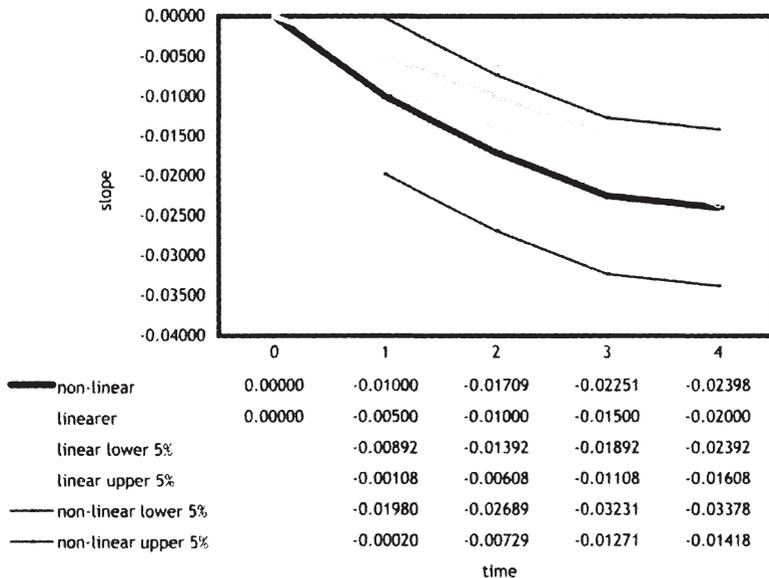


Diagramm 123: Linearer vs. nicht-linearer Verlauf mit Konfidenzintervallen

4.3.6.10. Indikatorspezifisch zeitübergreifende diachrone Korrelationen - Kinodaten

Um die Modelle weiter zu verbessern, werden indikatorspezifisch zeitübergreifende diachrone Korrelationen zwischen den Residualvarianzen der beobachteten Werte in Diagramm 124 eingeführt.

Indikatorspezifisch deshalb, weil lediglich Korrelationen innerhalb einer Variablen möglich sind, nicht jedoch zwischen unterschiedlichen Variablen. Zeitübergreifend bzw. diachron sind Korrelationen zwischen unterschiedlichen Zeitpunkten innerhalb einer Variablen möglich. Die freie Schätzung der Residuenkorrelationen wird dadurch begründet, dass von unterschiedlichen Eigenschaftsebenen bzw. unterschiedlichen Dimensionen, welche auf die Variablen wirken, ausgegangen wird. Jene Residuen einer Variablen, welche nicht durch die latente Dimension zu einem Zeitpunkt erklärt werden, werden auch in den darauffolgenden Wellen nicht durch die zugrundeliegende latente Dimension erklärt. Werden Korrelationen zwischen Residualvarianzen erlaubt, wird im Modell Freiraum für abweichende Dimensionswirkungen geschaffen, welche nicht in der zugrunde gelegten latenten Dimension enthalten sind. Die Konsequenz des geschaffenen Freiraumes hat Verbesserungen der Fit-Indizes zur Folge und sollte mit Vorsicht genossen werden.



Diagramm 124: Modell mit indikatorspezifisch zeitübergreifenden diachronen Korrelationen

Der Ergebnisvergleich der Wachstumskoeffizienten der linear fixierten und frei geschätzten Modelle über fünf Wellen mit Probandenausschluss bei Fehlerwerten wird in Tabelle 69 ausgegeben. Die Erweiterung um die frei geschätzten Residuenkorrelationen bringt keine Verbesserungen der Signifikanzwerte der Wachstumsparameter. Es zeigt sich sogar eine leichte Verschlechterung.

4.3.6.11. Modellfitveränderungen - Kinodaten

Wenn die Wachstumsparameter linear geschätzt werden zeigen die Fit-Werte deutliche Verbesserungen in Tabelle 70 für ‚im Kino‘ und ‚vor dem

Kinobesuch'. Der Grund dafür liegt in der nicht wiederholten Bestrafung der nicht erklärbaren Residuenkorrelationen.

Wachstums- koeffizienten	Probanden- anzahl	Ohne Residuenkorrelationen		Mit Residuenkorrelationen	
		Linear	Frei	Linear	Frei
Im Kinosaal	1.220	-0,005 (-1,457)	-	-	-
Im Kino	695	0,010 (1,844)	-	0,009 (1,680)	0,017 (1,059)
Vor dem Kinobesuch	892	-0,005 (-2,151)	-0,010 (-1,945)	-0,005 (-1,902)	-0,010 (-1,701)

Tabelle 69: Linear vs. nicht-linear mit bzw. ohne indicatorspezifisch zeitübergreifende diachrone Korrelationen

Residuenkorrelationen	Ohne		Mit	
	Im Kino	Vor dem Kinobesuch	Im Kino	Vor dem Kinobesuch
Chi-Square Test of Model Fit: Value	2494,289	6445,045	249,522	488,205
Degrees of Freedom	194	307	154	257
P-Value	0,0000	0,0000	0,0000	0,0000
CFI	0,675	0,479	0,986	0,980
TLI	0,681	0,491	0,983	0,977
Information Criteria:				
Number of Free Parameters	36	43	76	93
Akaike (AIC)	34643,213	48242,709	32478,446	42385,869
Bayesian (BIC)	34806,793	48448,828	32823,783	42831,662
Sample-size adjusted BIC ($n^*=(n+2)/24$)	34692,487	48312,268	32582,469	42536,311
RMSEA:				
Estimate				
90 Percent C.I.	0,131	0,150	0,030	0,032
Probability RMSEA <= .05	0,126-0,135 0,000	0,147-0,153 0,000	0,023-0,037 1,000	0,027-0,036 1,000
SRMR	0,088	0,128	0,030	0,042

Tabelle 70: Fit - Linear vs. nicht-linear mit bzw. ohne indicatorspezifisch zeitübergreifende diachrone Korrelationen

4.3.6.12. Einführung von Zeitpunktparametern - Kinodaten

Im folgenden Modell werden die Zeitpunkte der Beobachtung berücksichtigt, welche bis dato zwischen den fünf Bewertungen als konstant betrachtet wurden. Dies resultiert aus den fehlenden Kinostandortbewertungen zu einem oder mehreren Zeitpunkten oder durch die Bewertung eines anderen Kinos in dazwischen liegenden Wellen. Der nun verwendete Datensatz beinhaltet lediglich Fälle, in welchen über fünf Beobachtungen, egal zu welchem Zeitpunkt diese gemacht wurden, ein identischer Kinostandort bewertet wurde. Zusätzlich werden hier die Zeitpunkte als eigene Variable in das Modell eingebracht.

Es sind zwei Varianten der Zeitpunktvariablenbestimmung denkbar, sprich wie die Zeitabstände in das Modell einfließen. Die erste Möglichkeit betrachtet lediglich jene Monate als zu berücksichtigendes Zeitintervall, in welchen auch tatsächlich ein Kino bewertet wurde. Hier gilt die Annahme, dass die Häufigkeit der Bewertung eines x-beliebigen Kinos die Antworten in der folgenden Welle beeinflusst. Dies kann aufgrund der verstärkten geistigen Auseinandersetzung des Probanden mit dem Thema Kinostandort herrühren und kommt in diesem Modell verstärkt zur Geltung. Die zweite Möglichkeit berücksichtigt alle Zeitpunkte, gleichgültig ob in diesen ein anderer Kinostandort bewertet wurde oder nicht. In der folgenden Analyse wird letztere Variante getestet und alle verstrichenen Monate zwischen zwei Messungen berücksichtigt. Probanden mit fehlenden Werten werden aus dem Datensatz entfernt. Aufgrund von Schätzproblemen, ausgehend von den ermittelten Startwerten des Modells, werden die Items verglichen mit den vorangegangenen Modellen in entgegengesetzter Richtung codiert. Die Vorzeichen der Ergebnisse müssen für die Skala ‚sehr zufrieden (1)‘ bis ‚nicht zufrieden (6)‘ interpretiert werden und nicht wie in den Ergebnissen zuvor mit ‚sehr zufrieden (6)‘ bis ‚nicht zufrieden (1)‘. In Tabelle 71 ist erkennbar, dass es im linearen Modell über fünf Wiederholungen keine Signifikanzverbesserungen der Wachstumsparameter gibt. Es wird angenommen, dass derartig kurze Zeitabstände, welche sich innerhalb weniger Monatsintervalle befinden, in der Bewertung eines Kinostandortes keine bzw. eine zu geringe Rolle spielen.

In einer letzten Erweiterung wird über fünf Wellen ein linearer Wachstumsparameter mit erlaubten Residuenkorrelationen geschätzt. Zusätzlich werden lediglich jene Bewertungen der Probanden verwendet, welche nur einen speziellen Standort betreffen. Somit wird auch der verzerrende Einfluss unterschiedlicher Standorte ausgegrenzt. Die Zeitabstände werden wie in obigem Modell berücksichtigt und Personen mit fehlenden Werten nicht ersetzt. Es wird

wiederum die Skala ‚sehr zufrieden (1)‘ bis ‚nicht zufrieden (6)‘ verwendet. Auch für einen einzelnen Standort sind Wachstumsparameter nachweisbar. ‚Vor dem Kinobesuch‘ weist ohne Berücksichtigung der Zeitabstände einen signifikanten Wachstumsparameter auf. Daraus lässt sich schlussfolgern, dass für einen speziellen Kinostandort über die Zeit hinweg Veränderungen in der Zufriedenheit nachweisbar sind. Diese zeigen einen negativen Trend. Die Probanden bewerten über fünf Bewertungen hinweg den Kinostandort #18 schlechter in Bezug auf ihre Zufriedenheit.

	Probanden	Ohne Residuenkorrelationen unter Berücksichtigung der Zeitabstände	Mit Residuenkorrelationen unter Berücksichtigung der Zeitabstände
Im Kinosaal	1.220	0.002 (1.301)	-
Im Kino	695	-0.005 (-1.705)	-0.005 (-1.544)
Vor dem Kinobesuch	892	0.003 (2.132)	0.003 (1.890)

Tabelle 71: Berücksichtigung der Zeitabstände

Kino #18:	Probanden	Mit Residuenkorrelationen ohne Berücksichtigung der Zeitabstände	Mit Residuenkorrelationen unter Berücksichtigung der Zeitabstände
Im Kinosaal	282	0.016 (1.577)	-
Im Kino	182	-	0.003 (0.580)
Vor dem Kinobesuch	223	0.007 (2.098)	0.003 (1.687)

Tabelle 72: Berücksichtigung der Zeitabstände und Standorte

4.3.6.13. Zusammenhang zwischen *intercept* und *slope* - Kinodaten

Diagramm 125 stellt *scatterplots* dar, wobei die personenspezifischen *intercepts* und *slopes* für die drei latenten Dimensionen gegeneinander geplottet sind. Das den Ergebnissen zugrunde liegende Modell beschreibt ein Wachstum über fünf Wellen mit linearem Wachstumsparameter. Nimmt der Korrelationskoeffizient negative Werte an, weisen jene Probanden, welche mit einem höheren Zufriedenheits-*intercept* starten, einen stärkeren Abfall der Zufriedenheitswerte auf, im Vergleich zu Personen, welche auf einem niedrigeren Zufriedenheitslevel starten. Nimmt der Koeffizient positive Werte an, weisen jene Probanden, welche auf einem höheren Zufriedenheitslevel starten, ein stärkeres Wachstum auf, verglichen mit Personen auf einem niedrigeren Startwert des Zufriedenheitslevels. Die *intercept-slope*-Korrelation der Dimension ‚im Kinosaal‘

ist signifikant mit einem Korrelationswert von $-0,007$ ($-3,236$) und in der obersten Graphik dargestellt. ‚Im Kino‘ mit $0,004$ ($0,970$) und ‚vor dem Kinobesuch‘ mit $0,001$ ($1,102$) sind nicht signifikant. Bei den unteren beiden Graphiken muss die y-Achse im Vergleich zur x-Achse künstlich in die Länge gezogen werden, um die Korrelationsrichtung visualisierbar zu machen. In der ersten Graphik stimmen die Skalenabstände visuell überein.

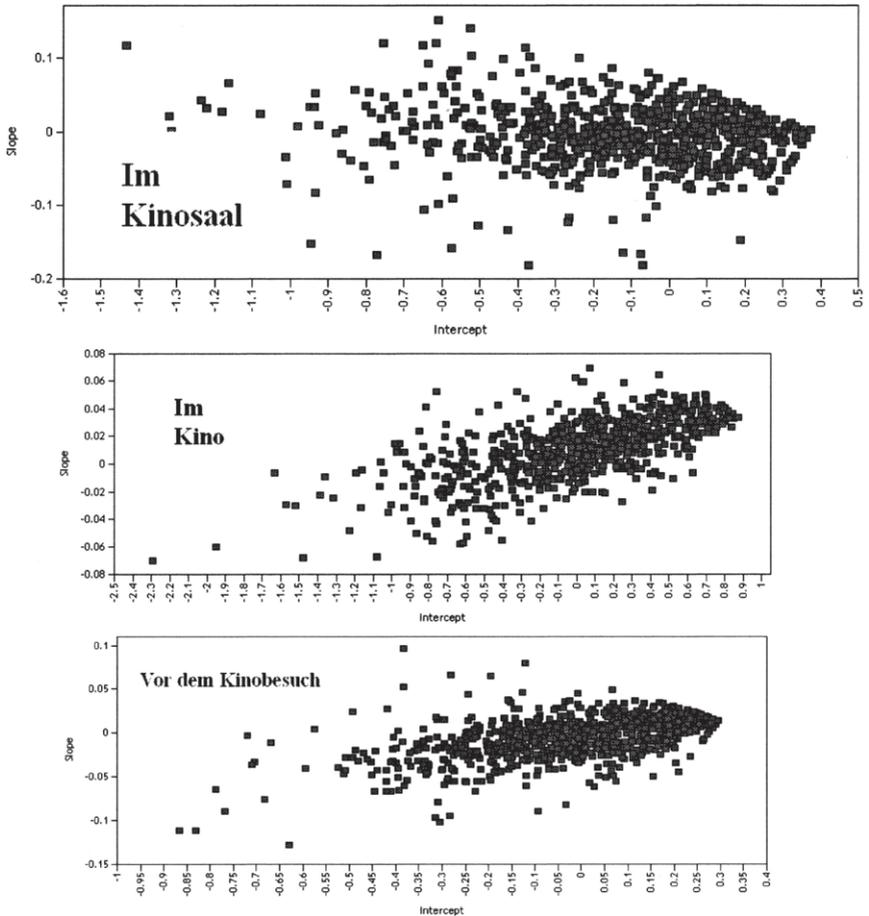


Diagramm 125: Intercept-Slope-Korrelationen

Im Folgenden geht es darum, Subpopulationen unterschiedlicher Wachstums-trajektorien zu bestimmen. Der Aufbau des restlichen Kapitels folgt einer chronologischen Abhandlung der Modellentwicklung.

4.3.6.14. A-posteriori ermittelte Populationsunterschiede

Nagin (1999) verwendet den *semi-parametric group-based approach* (SPGB-Ansatz). Zur Analyse individueller Entwicklungstrajektorien gibt es zwei Ansätze, das hierarchische Modell und die *latent curve*-Analyse. Beide berechnen, *unconditional* und *conditional*, die Populationsverteilung des Wachstumsparametermittelwertes und Varianzen des Wachstums über die Population. Dabei werden Populationsverteilungen des Wachstums als multivariate kontinuierliche Verteilungsfunktionen angesehen. Sie sind für die Identifizierung unterschiedlicher Kategorien von Clustertrajektorien nicht geschaffen, sondern eher, um zu beschreiben, wie Wachstumspatterns kontinuierlich über die Population verteilt sind. Der SPGB-Ansatz identifiziert unterschiedliche Cluster individueller Trajektorien innerhalb einer Population um Profile dieser zu erstellen und nutzt dafür Mischungen passender Wahrscheinlichkeitsverteilungen. Er geht von der Zusammensetzung der Population aus unterschiedlichen *mixtures* unterschiedlicher Gruppen, welche über ihre Entwicklungstrajektorien definiert sind, aus. Der *group-based* Ansatz nutzt eine multinomiale Modellstrategie um homogene Cluster von Entwicklungstrajektorien zu identifizieren. Er dient zur Überprüfung des Vorhandenseins von aus der Theorie vorhergesagten in der Population vorhandenen Wachstumstrajektorien.

Taxonomische Theorien sagen unterschiedliche Trajektorien von Entwicklungen über Subpopulationen voraus. Dabei geht es jedoch vorwiegend um die Gründe und Konsequenzen unterschiedlicher Entwicklungstrajektorien, als um den Vorschlag unterschiedlicher Gruppen. Entwicklungsbedingte Taxonomien stellen somit diagnostische Klassifikationen dar. Im Vergleich dazu bestätigen prototypische Klassifizierungen beispielsweise Unschärfe in der Klassifikation und Variation innerhalb der Gruppe. Sie dienen zur Vervollständigung von Theorien, welche prototypische Entwicklungen innerhalb einer Population vorhersagen. Es wird also auf die Unsicherheit der Gruppenzugehörigkeit Rücksicht genommen. Daher sind keine Faktoren notwendig, welche die Gruppenzugehörigkeit bestimmen. Es können Fragen überprüft werden, welche das Vorhandensein von in der Theorie vorhergesagten homogenen Entwicklungstrajektorien überprüfen. Man erspart sich die Zuordnungsregeln anhand von subjektiven Kriterien um Kategorien von Entwicklungstrajektorien zu schaffen. Diese können zwar nützlich sein, unterliegen aber Einschränkungen. Existenz von unterschiedlichen Entwicklungstrajektorien, welche einer taxonomischen Theorie unterliegen, können nicht getestet werden. Sie müssen vorweg angenommen werden, sonst führt z.B. *overfitting* zu Trajektoriegruppen,

welche lediglich aus Zufallsvariation entstehen. Weiters liefern zuvor spezifizierte Regeln keine Basis für die Präzisierung der Kalibrierung von individuellen Klassifikationen der verschiedenen Gruppen, welche die Taxonomie ausmachen. Die SPGB-Methode dient zur Bestimmung der Anzahl an Gruppen, welche am besten zu den Daten passt und liefert zusätzlich die *posterior*-Wahrscheinlichkeit der Gruppenzugehörigkeit um die Präzision der Gruppenzuweisung zu evaluieren. Zusammengefasst folgen hierarchische und *latent curve*-Modelle einer kontinuierlichen Populationsverteilung der Entwicklung, um durchschnittliche Entwicklungstendenzen und deren Variabilität zu schätzen und durch Kovariate zu erklären. *Mixture*-Modelle dienen zur Identifizierung unterschiedlicher prototypischer Wachstumsverläufe innerhalb der Population, um die Wahrscheinlichkeit der Population zu bestimmen, mit welcher diese diesem Verlauf folgt und diesen mit Kovariaten zu erklären. Der Grund liegt darin, dass Entwicklungen nicht regelmäßig über die Population variieren, sondern Gruppen großen Unterschieden und unterschiedlichen Verläufen und Einflüssen unterliegen.

Anhand von zwei Datensätzen stellt Nagin (1999) Lösungen für das Poisson-basierte Modell mit Zähldaten (Poisson Verteilung), für das *censored normal*-Modell für psychometrische Skalen mit Clustern am Skalenminimum oder – maximum oder beiden (*censored*-Normalverteilung) und für das binäre Modell (*binary logit*-Verteilung) für die quadratische Trajektorielösung, vor. Wäre die Gruppenzugehörigkeit bekannt, könnten die Trajektorieparameter über Poisson, *censored normal (tobit)* bzw. *logit*-Regression berechnet werden. Für vorliegendes Modell werden diese jedoch aus den Daten heraus identifiziert. Zur Modellselektion ist LRT zur Gruppenanzahlbestimmung nicht geeignet. Die Nullhypothese, drei Komponenten vs. mehr als drei Komponenten, liegt am Rande des Parameterraumes und die klassischen asymptotischen Ergebnisse welche einem LRT unterliegen können nicht gehalten werden. Außerdem sind Modelle verschiedener Gruppenanzahl nicht genestet und für den LRT somit nicht geeignet. Deshalb wird BIC empfohlen, welcher für genestete und nicht genestete Modelle geeignet ist. Nagin (1999) beschreibt diesen Umstand. Eine Gruppe kann überflüssig sein indem der Anteil der Gruppe null ist, andererseits können die drei Parameter, welche die Trajektorie einer Gruppe beschreiben, in die einer anderen zusammenfallen. Die passende Anzahl an Parametern, ob einer oder drei, ist nicht mehr bestimmbar. BIC ist immer negativ und jenes Modell mit dem geringst negativen Wert zu bevorzugen. Es kann jedoch keine Aussage über die Quantität der BIC-Verbesserung gemacht werden. Für den Bayes-Faktor, welcher die *odds* zweier vergleichender Modelle berechnet, also das Verhältnis der Wahrscheinlichkeit, dass *i* das korrekte Modell ist, verglichen mit

Modell j , gibt es eine Annäherung. Ein Ergebnis aus Formel 24 für zwei Modelle von 1 bedeutet gleich gut, >10 Bevorzugung Modell i und $<1/10$ Bevorzugung Modell j .

$$e^{BIC_i - BIC_j}$$

Formel 24: Bayes-Faktor-Annäherung (zwei Modelle)

Formel 25 zeigt das Maß zum Vergleich von mehr als zwei Modellen. p_j ist die *posterior*-Wahrscheinlichkeit dass Modell j das richtige ist. Je näher bei eins, desto besser der Wert.

$$p_j = \frac{e^{BIC_j - BIC^{\max}}}{\sum_j e^{BIC_j - BIC^{\max}}}$$

Formel 25: Bayes Faktor Annäherung (mehr als zwei Modelle)

Für drei Gruppen gibt es $3^4=81$ Möglichkeiten an Trajektorien und somit 256 Möglichkeiten. Auswahlprozesse wurden bereits oben beschrieben. Die Einbringung von Kovariaten gleicht die Wahrscheinlichkeiten der Gruppenzugehörigkeit unter dem Einfluss des Levels weiterer Faktoren an. Nagin (1999) zeigt auf, dass die Profile der Gruppen und die Identifizierung probabilistisch und nicht sicher sind. F - und χ^2 -basierte Tests fallen aus, da diese keine Klassifizierungsfehler der Gruppen annehmen. Die geschätzten Koeffizienten sind als Einfluss der Kovariate auf die Gruppenzugehörigkeitswahrscheinlichkeit relativ zur Kontrastgruppe zu interpretieren. So können Kovariate die Wahrscheinlichkeit der Zugehörigkeit zu einer Gruppe steigern oder senken. Die Auswahl der relevanten Kovariate führt über die Schätzung des am besten brauchbaren Modells. Dann werden entweder über multinomiale *logit*-Modelle die Gruppenzugehörigkeit mit den individuellen Faktorenlevels verknüpft oder die Gruppenzugehörigkeitswahrscheinlichkeiten auf die individuellen Levels der Faktoren regressiert. Die beiden Möglichkeiten tendieren jedoch dahin mehr Kovariate einzubringen, da sie die Unsicherheit der Gruppenzugehörigkeit nicht berücksichtigen.

Muthén und Shedden (1999) stellen eine erweiterte Form der *finite mixture models* (FMM) vor, wobei die latenten Klassen, also die Klassenzugehörigkeit zu latenten Wachstumstrajektorien, die Wahrscheinlichkeit weiterer beobachteter Variablen beeinflussen. *Random coefficient modeling* mit *mixtures* wiederholter Messungen vermeidet somit die Annahme normalverteilter Zufallskoeffizienten, *random coefficients*. Es werden EM mit ML-Schätzung und *bootstrap*-

Standardfehlern verwendet und die E- und M-steps detailliert besprochen. Veranschaulicht wird das Beispiel anhand einer Alkoholismusstudie, wobei der wiederholt beobachtete Alkoholkonsum, also der Entwicklungsverlauf des Alkoholgenusses, die spätere beobachtete Wahrscheinlichkeit der Alkoholabhängigkeit vorhersagt. Dies kann durch logistische Regression der späteren Abhängigkeit der *intercept*- und *slope*-Koeffizienten erreicht werden.

Sie beschreiben das innewohnende Problem. Ein Wachstumskoeffizient hat, je nachdem welche Werte die anderen Koeffizienten annehmen, eine unterschiedliche Bedeutung. Um die Vorhersagekraft des Kurvenverlaufes besser darzustellen, schlagen sie deshalb *extended FMM* vor. Dabei werden durch gemeinsame Schätzung unterschiedliche Kurvenverläufe durch klassenvariierende *random coefficient means* aufgefangen und durch logistische Regression die Alkoholabhängigkeit dieser Klassen miteinbezogen. Würde das Modell nicht auf einmal geschätzt werden, bräuchte man dafür drei Schritte. Zuerst müsste das *finite mixture growth*-Modell geschätzt werden. Dann müsste die individuelle höchstwahrscheinliche Klassenzugehörigkeit basierend auf den *posterior*-Wahrscheinlichkeiten für die Klassen, welche vom geschätzten Modell bekannt sind, geschätzt werden und im Anschluss die Alkoholabhängigkeit auf die geschätzte Klassenzugehörigkeit regressiert werden. Fehler passieren aufgrund der Zuweisung eines Individuums in eine einzige Klasse wobei die *posterior*-Wahrscheinlichkeiten normalerweise aber niemals null für eine Klasse sind. Somit beinhaltet der *extended FMM*-Ansatz zwei Verbesserungen. Zum einen kann die Klassenzugehörigkeitswahrscheinlichkeit als Funktion der Kovariate variieren. Zum anderen werden die Wachstumsparameter jeder Klasse von den Kovariaten beeinflusst. Zusätzlich kann die latente Klassenvariable mehr als eine binäre *outcome*-Variable beeinflussen. Das Modell besteht aus kontinuierlich über die Zeit hinweg beobachteten Variablen, den binären *outcome*-Variablen, Kovariaten, latenten kontinuierlichen Wachstumsvariablen, der kategorialen Klassenvariable mit multinomialer Verteilung und den Residuen, welche unkorreliert mit den anderen Variablen im Modell normalverteilt mit Mittelwert null sind. Die kontinuierlichen Wachstumsvariablen werden mit der latenten kategorialen Klassenvariable und den beobachteten Kovariaten verbunden. Daraus entstehen wiederum normalverteilte Residuen, welche zu den anderen Variablen nicht korreliert sind und Mittelwert null aufweisen. Die Mischkomponente, also die latente Klassenvariable, wird mit den Kovariaten durch multinomiale *logit*-Regression für ungeordnete polytome Antwortvariablen in Beziehung gesetzt. *Finite mixture* entsteht, da die bedingte Verteilung der beobachteten wiederholten Variablen und den *outcome*-Variablen, gegeben die Kovariate, durch Parameter aufgefangen wird, welche über die

Kategorien der Klassenvariablen variieren. Der Mittelwertvektor der wiederholt beobachteten Variablen variiert durch die Verbindung mit der latenten Klassenvariablen und die Wahrscheinlichkeit der *outcome*-Variablen variiert durch Verbindung mit der latenten Klassenvariablen. Somit liegt die Generalisierung des Modells in der Einbindung direkter Effekte der Kovariate auf die beobachteten wiederholten Variablen, die *time-varying*-Kovariate und Effekte der Kovariate auf die latente Klassenvariable. Daraus folgen klassenspezifische Effekte. Eine strenge Annahme ist, dass die *outcome*-Variablen von den Kovariaten rein über die latente Klassenvariable abhängen. Eine Alternative dazu wären direkte Effekte der Kovariate auf die *outcome*-Variablen. Die Wahrscheinlichkeiten der *outcome*-Variablen variieren nämlich vielleicht nicht nur als Funktion der Klassenzugehörigkeit, sondern auch als Funktion der Kovariatencharakteristik für die gegebene Klassenzugehörigkeit.

Li et al. (2001) geben ein Beispiel für *general growth mixture models* (GGMM), welche die früheren *random coefficient growth models/latent curve*-Analyse um eine kategoriale latente Trajektorievariable mit multinomialer Verteilung, welche latente Klassen repräsentiert, im Kontext latenter Klassenmodelle erweitert. Die Generalisierung von *latent growth models* (LGM) zu *finite-mixture latent trajectory*-Modellen für die GGMM stammt von Muthén (2001). Sie beinhaltet die Variation der *random growth effects conditional* zu Kovariaten. Zusätzlich können unterschiedliche latente Klassen unterschiedliche voraussagende Beziehungen zwischen Kovariaten und latenten Wachstumstrajektorien aufweisen. Die *mixture*-Komponente wird über eine multinomiale logit-Regression mit den Kovariaten, ungeordnete polytome Antworten, zusammenhängt. Der Artikel behandelt FGMM, wobei unterschiedliche Klassen unterschiedliche *random coefficients* bezüglich der Faktormittelwerte, deren Varianz und Kovarianzstruktur und zeitspezifische Fehlervarianzen aufweisen. FMM kommt durch die *conditional* Verteilung der kontinuierlich beobachteten Variablen und der *outcome* Variablen, gegeben die Kovariaten und variiert durch die latente kategoriale Klassenvariable. In *growth mixture models* (GMM) ist, im Vergleich zu *multiple-group latent curve analysis*, die Gruppenzugehörigkeit unbekannt und wird aus den Daten abgeleitet. Ein weiterer Vorteil ist die Ablegung der Annahme einer einzigen Population, wodurch Kovariate unterschiedliche Einflüsse auf unterschiedliche Gruppen und Gruppen unterschiedliche Auswirkungen auf die *distal outcomes* haben können. *Distal outcomes* stehen für die Resultate aufgrund von Wirkungseffekten.

χ^2 -basierte Statistiken wie LRT werden für genestete Modelle verwendet. Modelle mit unterschiedlicher Clusteranzahl sind nicht genestet und die

Ordnungsbedingungen für die Gültigkeit der asymptotischen Verteilung des LRT sind in *mixture*-Modellen verletzt. AIC, BIC oder ABIC, wobei letzterer bevorzugt wird, werden verwendet. Sie alle bestrafen Modelle mit mehr Parametern. Nach Schätzung der Parameter werden die *posterior*-Wahrscheinlichkeiten für die Zugehörigkeit zu einer bestimmten Klasse für die Individuen berechnet. Diese werden auch zur Erzeugung der gewichteten Stichprobenmittelwertvektoren und der Kovarianzmatrizen für jede Klasse verwendet. Li et al. (2001) untersuchen den Alkoholkonsum über die Zeit, beeinflusst durch die Kovariaten Geschlecht, Familienstatus, abweichendes Verhalten und die Beziehung zu Gleichgesinnten mit abweichendem Verhalten. Zusätzlich wird der zukünftige Alkohol/Zigaretten/Marijuana-Gebrauch in späteren Jahren beobachtet und als *distal outcome* verwendet. Zuerst werden ein *unconditional*-Einklassenmodell, dann *unconditional*-Mehrklassenmodelle geschätzt, über Informationskriterien verglichen und die bevorzugte Klassenanzahl bestimmt. Dann werden zwischen den genesteten Modellen mit klassenvariierenden Mittelwerten der *intercept*- und *slope*-Faktoren, mit klassenvariierenden Varianzen und Kovarianzen der *intercept* und *slope*-Faktoren und mit klassenvariierenden zeitspezifischen Fehlervarianzen, über den χ^2 -Differenztest der LRT-*statistic* Modelle mit und ohne Einschränkungen bezüglich der Invarianzen verglichen. Dann werden nichtsignifikante *slopes* und deren Kovarianzen zum *intercept* auf null fixiert. Um unterschiedliche Einflussstärken der Kovariaten zu den Wachstumsstrukturen zu bestimmen wird ein *conditional* GMM über Regression zwischen diesen gerechnet, welches Variation in den klassenspezifischen Wachstumsfaktoren auffängt. Verglichen wird das Einklassenmodell mit Kovariaten und das Mehrklassenmodell mit den Kovariaten über *log likelihood*-Wert, AIC, BIC und ABIC. Dann wird das Modell um *distal outcomes* für die jeweiligen Wachstumstrajektorien durch Regression erweitert, um die Wahrscheinlichkeit in den *outcome*-Variablen über die latenten Trajektorieklassen variieren zu lassen. Dann wird noch die Verbindung der Kovariaten mit den *distal outcomes* über logistische Regression eingeführt, um zu bestimmen, ob die Wahrscheinlichkeiten in den *distal outcomes* als eine Funktion der Kovariate für eine gegebene latente Klassentrajektoriezugehörigkeit variiert und über LRT verglichen. Es werden unterschiedliche Wachstumstrajektorien mit klassenvariierenden Zufalls-koeffizienten der Mittelwerte und Varianzen, mit variierenden Klassen-zugehörigkeitswahrscheinlichkeiten als eine Funktion der Kovariate, und die logistische Regression der *distal outcomes* auf die Klassentrajektorie gleichzeitig geschätzt. Li et al. (2001) beschreiben den Unterschied zum SPGB *mixture model*-Ansatz. Dieser hat keine Annahmen bezüglich der Verteilung der Wachstumsparameter und somit wird eine unspezifizierte möglicherweise

kontinuierliche Verteilung der Populationsheterogenität in Entwicklungstrajektorien approximiert. Muthéns Ansatz geht von Normalverteilung in der Population aus. Li et al. (2001) weisen weiters auf *multiple-cohort growth mixture*-Analyse hin, welche als *multiple-group single-population modeling framework* angesehen werden kann, um beispielsweise Fehlerwertprobleme zu lösen.

Duncan et al. (2002) beschreiben anhand einer Alkoholismusstudie den schrittweisen Aufbau eines komplexen FMM, welches Heterogenität innerhalb der Stichprobe annimmt und durch eine kategoriale Klassenvariable beschrieben wird. *Latent variable mixture modeling* besteht erstens aus der *mixture portion*, welche den Einfluss der kategorial latenten Variable auf die binären/geordneten kategorialen/kontinuierlichen beobachteten latenten Klassenindikatoren beschreibt und Regressionen auf andere Variablen oder Kovariate misst. GMM bezieht zusätzlich wiederholte Beobachtungen mit ein. Zweiter Teil ist die *structural portion*, wobei alle Parameter über die latenten Klassen variieren und die kontinuierlichen latenten Variablen auch kontinuierliche *outcome*-Variablen, aber nicht kategoriale, beeinflussen können (Duncan et al., 2002). Eine Annahme, welche nicht immer rechtfertigbar ist, ist die Gleichheit des Mess- und Strukturmodells über alle latenten Klassen. Zuerst wird in genanntem Aufsatz ein *single-class growth curve*-Modell, ohne Prädiktoren, also *unconditional*, gerechnet. *Conditional* soll immer erst nach den Tests der *unconditional*-Analysen eingeführt werden. Werden im *single-class*-Modell signifikante Varianzen des latenten *slopes* und/oder *intercepts* gefunden, wird zum *two-class mixture*-Modell, auch *multiple-class growth curve*-Modell genannt, übergegangen. Später werden Kovariate auf die latenten Wachstumsfaktoren und die latenten Klassen eingeführt, *conditional*. Durch die eingeführten Regressionen sollten die BIC-Werte durch besseren Modellfit sinken. Weiters wird eine Regression zwischen der latenten Klassenvariable und einem *mixture*-Indikator, einer *outcome*-Variablen, eingeführt, später auch zwischen Prädiktor und *outcome*-Variablen. Begonnen wird immer mit einem GMM mit kontinuierlichen latenten Wachstumsvariablen und einer latenten kategorialen Klassenvariable. Darauf folgend werden Prädiktorvariablen eingeführt. FMM ergibt sich dadurch, da die bedingten Verteilungen, welche durch die Beziehung des Prädiktors und der beobachteten Variablen und der *outcome*-Variablen entstehen, über die unterschiedlichen Klassen variieren.

Die latente Klassentheorie geht von einer statischen sich nicht verändernden diskontinuierlichen latenten Variablen aus, welche die Population in latente Klassen unterteilt. Natürlich sind auch dynamische latente Variable möglich. Im

Unterschied zu statischen latenten Variablen ändern sich die dynamischen über die Zeit hinweg, beispielsweise in *latent class models for dynamic processes*, nämlich *latent transition analysis* (LTA). Es können *latent class*- und *latent transition*-Modelle definiert werden, indem dynamische latente Variable eine systematische Veränderung über die Zeit aufweisen, *stage-sequential development*. Mischmodelle im Kontext von Kovarianzstrukturmodellen gehören zu den SEMs. Der GMM-Ansatz verbindet kategoriale und kontinuierliche latente Variable in einem Modell. Die Generalisierung von LGM, nämlich die in Richtung der *finite-mixture latent trajectory*-Modelle, auch GGMM genannt, welche zusätzlich zur unterschiedlichen Variabilität um den Mittelwert innerhalb der Gruppen wie im FMM, gleichzeitig auch logistische Regressionen von *outcome*-Variablen auf die latente Klasse schätzen. Diese Modelle ermöglichen die Erweiterung um variierende Klassenzugehörigkeitswahrscheinlichkeiten als Funktion eines Sets an Kovariaten, beispielsweise durch den Einfluss der Kovariaten in jeder Klasse auf die Werte der latenten Wachstumsparameter, und um *outcomes* der latenten Klassenvariable (Duncan et al., 2002). *Mixture multivariate growth*-Modelle erlauben simultane Wachstumsprozesse durch *multiple-outcome growth mixture*-Modelle um multivariate Veränderungen zu untersuchen. *Multigroup growth mixture*-Modelle erlauben unterschiedliche Gruppen mit unterschiedlichen latenten Wachstumsstrukturen unabhängiger Stichproben, wobei Modelle über bekannte Gruppen mit Einschränkungen über diese geschätzt werden.

Schaeffer et al. (2003) verwenden ein GGMM um die Entwicklung frühzeitig aggressiven Verhaltens mit später auftretendem antisozialem Benehmen zu verbinden. Daten werden von der ersten bis zur siebenten Schulstufe und im Alter 19 bis 20 erhoben, und zwar 36 Items aus dem *Teacher Observation of Classroom Adaptation – Revised* (TOCA-R) wobei die Lehrer aggressives und abweichendes Verhalten, Aufmerksamkeits- und Konzentrationsprobleme und Gruppenablehnung, beurteilten. Kovariate stellen Items aus dem *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), jugendliche und erwachsene Gerichtsbeschlussaufzeichnungen und die Berechtigung für Gratisessen anstatt des Familieneinkommens, dar.

Die Parameter werden mit FIML unter MAR-Annahme für alle verfügbaren Zeitpunkte geschätzt. MAR geht von Zufälligkeit der fehlenden Daten aus. Dabei besteht kein Zusammenhang zum interessierten *outcome*. Der MAR-Ansatz kann auch gehalten werden, wenn nach Einflussbereinigung anderer Variablen, dieser als zufällig angenommen werden kann. Um das Ausmaß der fehlenden Werte beurteilen zu können, liefert MPlus die *covariance coverage*-

Matrix, welche die Proportion der verfügbaren Beobachtungen jeder Indikatorvariablen und den Paaren ausgibt. Das Minimum an Abdeckung liegt bei 0.10. In der Studie werden Werte zwischen 0.40 und 0.91 als mehr als akzeptabel angesehen. Korrelationen zwischen Werten unterschiedlicher Zeitpunkte sind niedriger, je weiter die Zeitpunkte voneinander entfernt liegen. Die geschätzten und beobachteten Mittelwerte liegen nahe beieinander. GGMM umgehen die Annahmen der Invarianz zwischen den Klassen und Zeitpunkten und den unkorrigierten Schätzungen der Kovariaten und *distal outcomes*. Auch Schaeffer et al. (2003) erwähnen den bevorzugten personenzentrierten Fokus, um individuelle Entwicklungsdifferenzen zu schätzen, verglichen mit dem variablenzentrierten Fokus und die empirische anstatt der subjektiven Einteilung der Personen. Nachteile vorangegangener Studien sehen sie beim SPGB-Ansatz in der Annahme, dass die Entwicklungen für alle Individuen gemeinsam geschaffen werden, was bei unterschiedlichem Verhalten und der daraus folgenden Variabilität in den Klassen nicht angenommen werden darf. Weiters bemängeln sie das bisherige Fehlen der Verbindung zwischen den Heterogenitätslevels und Trajektorieklassen, den Risikofaktoren der Klassenzugehörigkeit und den *distal outcomes*. Ziele liegen in der Identifizierung von Trajektorien, späteres Verhalten vorherzusagen und Faktoren, welche die Trajektorien beeinflussen, zu bestimmen. GGMM liefert unverzerrte Schätzer der Verbindung zwischen Kovariaten und *distal outcomes* und Trajektorieklassen, durch den Abgleich der Schätzungen durch die Unsicherheit der Klassenzugehörigkeit. GGMM schätzen die latenten Variablen aus mehreren Indikatoren, wobei die multiplen Indikatoren der latenten Wachstumsparameter wiederholte univariate *outcomes* zu den unterschiedlichen Zeitpunkten darstellen. Im Gegensatz zur Annahme, dass die Population einer einzigen kontinuierlichen Verteilung folgt, untersuchen GGMM das Vorhandensein zweier oder mehrerer diskreter Klassen, mit Entwicklungen von Individuen durch optimale Zuordnung zu latenten Klassen. Zeigt das Modell mit mehreren latenten Klassen einen bessern Fit, kann dies nachgewiesen werden. GGMM ist dem SPGB-Ansatz ähnlich. Beide schätzen unterschiedliche Trends in den Klassen, GGMM hingegen erlaubt klassenspezifische Variation, was meist zu besserem Modellfit führt. Gibt es diese nicht, liefern beide Ansätze die gleichen Ergebnisse.

Schaeffer et al. (2003) vergleichen am Beginn *intercept-only* mit *intercept+linear* und *intercept+linear+quadratic*. Die Modelle sind genestet und ein χ^2 -Differenztest zeigt die Bevorzugung des nicht-linearen Verlaufes. Es folgt die Hinzufügung von Residualvarianzen in der ersten und der zweiten Klasse durch Kovarianzen, da die Beurteilung von den gleichen Lehrern abgegeben wurde.

Anschließend werden die Varianz des quadratischen Wachstumsfaktors und die Kovarianz zwischen *intercept*- und *slope*-Faktoren auf null gesetzt, da diese, wenn frei geschätzt, nicht signifikant und schwach sind. Dies führt zur Verbesserung des Modellfits des Einklassenmodells. Es folgt ein Vergleich der nicht genesteten Klassenanzahlvergleiche über BIC, AIC, *sample-size adjusted* BIC (SSABIC) und dem *entropy*-Kriterium. Vier Klassen werden ausgewählt und zur weiteren Fitverbesserung werden Differenzen der Residuenvarianzen zwischen den Klassen erlaubt. Der Wachstumsparameter der nicht aggressiven Gruppe wird auf null fixiert. Die Regressierung der Klassenzugehörigkeit auf die genannten Kovariate führt nach Fixierung der Möglichkeit des Gratisessens durch χ^2 -Differenztest zu verbessertem Modellfit. Die zusätzliche Variation dieser Effekte zwischen den Klassen ist nicht signifikant. Regression der vier *outcome*-Variablen auf die Klassenzugehörigkeit, einzeln durchgeführt, da diese hohe Kollinearität aufweisen, liefert zusätzliche Einsichten klassenunterschiedlicher Auswirkungen.

Delucchi et al. (2004) verwenden *latent class growth models* (LCGM), welche auf FMM basieren. Im Kontext der SEM werden diese über GMM verarbeitet, wobei sie Heterogenität innerhalb der Trajektorien erlauben, was in LCGM nicht erlaubt ist. Delucchi et al. gehen detaillierter auf LCGM als SPGB-Ansatz ein. Diese Modelle verwenden eine multinomiale Strategie um homogene Cluster individueller Trajektorien zu finden und Einflüsse auf diese zu testen. Den Vorteil gegenüber der *classify-then-analysis*-Strategie sehen sie in der meist falschen Annahme, dass die Zuteilung zu den Clustern ohne Fehler passiert, was natürlich zu Verzerrungen führen kann. Deshalb werden simultan die Gruppenzugehörigkeit und -differenzen geschätzt, welche die Unsicherheit der Gruppenzugehörigkeit durch die Schätzung der Standardfehler beim Testen der Differenzen mit einbezieht. Delucchi et al. (2004) untersuchen, ob es Gruppen mit unterschiedlichen Profilen gibt, welche Trajektorien diese aufweisen und ob es weitere beeinflussende Variable gibt, welche die Steigung der Trajektorien beeinflussen. Dafür werden zuerst LCGMs zur Bestimmung der Clusteranzahl anhand der Höhe an konsumiertem Alkohol geschätzt. Die Modelle unterschiedlicher Anzahl werden über den approximierten Bayes-Faktor verglichen und das sparsamste gewählt, wobei idealerweise kein Cluster unter 5% der Gesamtstichprobe enthält. Der approximierte Bayes-Faktor berechnet sich aus BIC minus zweimal die Veränderung zwischen den BIC Werten der Modelle. Im Anschluss wird ein multinomiales logistisches Regressionsmodell mit den bestimmten Clustern mit den *baseline*-Variablen (Alter, Geschlecht, Familienstand und -einkommen, Ethnizität, alkoholbezogene soziale Konsequenzen, Abhängigkeitssymptome und ob im Vorjahr die Alkohol-

abhängigkeit aufgetreten ist) als Kovariaten eingebracht. Die signifikanten werden im Gesamtmodell erneut geschätzt und die wiederum signifikanten Kovariaten verbleiben im Modell. Zugehörigkeiten zu Clustern können sich dadurch ändern. Im dritten Schritt werden *time-varying* Kovariate (Empfehlungen von Freunden, Gesundheitseinrichtungen etc. in Bezug auf den Alkoholkonsum, ob Kontakt zu Sozialhilfe etc. aufgenommen wurde, Größe der Suchtmittel verwendenden sozialen Bekanntschaft, Anzahl der Tage bei den anonymen Alkoholikern und Missbrauchsbehandlung) durch *repeated measures general linear*-Modelle mit *single multinomial logistic regression* eingebracht. Die signifikanten verbleiben im Modell und das Gesamtmodell wird geschätzt und evaluiert. Es finden sich fünf Gruppen, *early quitters*, *light/non-drinkers*, *gradual improvers*, *moderate drinkers* und *heavy drinkers*.

Ein Modell mit *mean imputation* zum Ersatz fehlender Werte liefert gleiche Ergebnisse. Delucchi et al. (2004) sehen Probleme beim Versuch kontinuierliche Klassen diskret zu schätzen, im Verlust von Daten falls MCAR vorliegt bzw. in der Verzerrung der latenten Struktur falls MCAR nicht vorliegt und in fehlenden Modellen aufgrund von Vorgehensweisen die eventuell sparsamer sind. Weiters stellt das Auffinden einer ML-Lösung durch die Sensitivität des Algorithmus aufgrund der Startwerte ein Problem dar und die Subjektivität sinkt aufgrund des BIC Vergleiches anstatt der Suche nach objektiveren Kriterien. Aufgrund des *semi-parametric* LCGM anstatt des *parametric* GMM ist die Modellselektion schwieriger. GMM ist brauchbar, da die *response*-Variable normalverteilt ist, die logarithmierte Menge an Alkohol. LCGM wird aufgrund nicht normalverteilter Variablen verwendet. GMM besitzt die Restriktion multivariat normalverteilter Daten. Außerdem besteht die Unsicherheit, relevante Daten zur Schätzung von Heterogenität innerhalb der Klassen zu besitzen. Das Modell ist jedoch besser als *mixed-effects*-ANOVA für longitudinale Daten.

Jung und Wickrama (2008) besprechen *latent class growth analysis* (LCGA) und GMM. Dabei geht es um die Klassifizierung homogener Individuen innerhalb einer Gesamtpopulation. Es wird angenommen, dass interindividuelle Unterschiede in intraindividuellen Veränderungen über die Zeit hinweg bestehen. In herkömmlichen Wachstumsmodellen, welche bereits gezeigt wurden, beschreibt ein einziger Wachstumsverlauf alle Individuen einer Population und auch Kovariate beeinflussen alle Individuen gleich. Um den Unterschied darzustellen, wird der Vergleich zwischen personen- und variablenzentrierten Analysen herangezogen (Jung & Wickrama, 2008). Variablenzentrierte Analysen beschreiben Zusammenhänge von Variablen, beispielsweise über Faktorenanalyse, Regression oder SEM mit Beziehungen zwischen

Variablen. Personen-zentrierte Analysen beschreiben Zusammenhänge zwischen Personen mit dem Ziel diese in homogene Gruppen zu teilen, beispielsweise über Clusteranalysen, *latent class analysis* (LCA) oder FMM. Der logische Aufbau kann über *multilevel random effects*-Modelle erklärt werden. Diese beinhalten eine Variation um den Mittelwert, welche durch die *random effects* beschrieben wird. GMM teilt die Personen in Klassen und geht von unterschiedlicher *unique variation* der *intercepts* und *slopes* innerhalb der Klassen aus. LCGA teilt die Personen in Klassen, unterstellt jedoch eine Variation von null und somit identische Werte für alle Personen. Letztgenanntes Verfahren dient meist zur Bestimmung von Gruppen vor Verwendung von GMM. Zuerst wird eine einzelne Gruppe geschätzt und diese dann mit einer höheren Anzahl verglichen.

Ein gebräuchlicher Schätzer für FMM ist ML mit iterativem EM-Verfahren, welcher das Modell mit dem höchsten *loglikelihood*-Wert auswählt. Im E-Schritt werden die bedingten Wahrscheinlichkeiten betrachtet, im M-Schritt die unbekannt Parameter durch Maximierung der bedingten Wahrscheinlichkeiten aus dem E-Schritt geschätzt. Aufgrund der Stichprobenverteilung und den in den *mixture*-Modellen liegenden Mischungen an Verteilungen kann es zu Konvergenzproblemen und lokalen Maxima kommen. Höhere Anzahl an zufälligen Startwerten, höhere Anzahl daraus optimierter Lösungen und höhere Anzahl an Iterationen zum Auffinden der höchsten *likelihood*-Werte der zu optimierenden Lösungen, schaffen dafür Abhilfe. Ein zusätzlicher Vergleich der optimierten Lösungen zwischen den zwei höchsten *likelihood*-Werten der Startwertlösung sollte zum gleichen Ergebnis kommen, wobei der beste *likelihood*-Wert zumindest zweimal auftauchen sollte. Dies kann den Verdacht einer lokalen Lösung entschärfen und den einer globalen Lösung erhärten.

Die Evaluierung des Modellfits eines Mischmodelles kann über die *log likelihood* vorgenommen werden. Der χ^2 -basierte LRT dient zum Vergleich genesteter Modelle, nicht jedoch von Modellen unterschiedlicher Klassenanzahl. Dafür stehen AIC, BIC oder der *sample-sized* ABIC zur Verfügung, wobei letzter zu bevorzugen ist. Auch die *posterior*-Wahrscheinlichkeit des Individuums, einer Klasse anzugehören, kann betrachtet werden. In den Zeilen stehen die Individuen, welche sich in der Klasse befinden und in den Spalten die durchschnittlichen bedingten Wahrscheinlichkeiten. Liegen die Werte der Hauptdiagonalen nahe bei eins und die der oberen und unteren Dreiecksmatrix nahe bei null, ist dies ein gutes Indiz für sinnvolle Klassen. Als weitere Entscheidungshilfe, wie viele Gruppen gewählt werden sollen dient BIC, *Lo, Mendell and Rubin likelihood ratio test* (LMR-LRT) oder *bootstrap likelihood ratio test* (BLRT). Die letzten beiden vergleichen die Lösung mit k und $k-1$

Klassen. Bringt die zusätzliche Klasse eine signifikante Modellverbesserung, wird das komplexere Modell bevorzugt. Der rechenaufwändige BLRT wird bevorzugt, dann BIC und dann ABIC. Jedoch auch Forschungsfrage, Sparsamkeit der Modelle, theoretische Rechtfertigung und Interpretierbarkeit sind zu beachten. Weitere Modellbewertungen stellen auftretende Konvergenzprobleme, Anzahl der Personen pro Klasse, posteriore Wahrscheinlichkeit der Zugehörigkeit zu einer Klasse, als auch der *entropy*-Wert dar. Das Problem betrifft vor allem die Auswahl der Komponentenanzahl. LRT kann nicht verwendet werden, da die χ^2 -Verteilung der LR-*statistic* nicht gewährleistet ist. Lösungen dazu bieten *parametric bootstrap*, um die *p*-Werte dem *generalized likelihood ratio* (GLR)-Test anzunähern, Informationskriterien oder der Bayes-Faktor. *Gaussian mixture* um Ausreißer von gleichen Mittelwerten zu finden passt nicht. AIC überschätzt die Anzahl an Komponenten, BIC unterschätzt diese. *Informational complexity criterion* (ICOMP) verhält sich besser als AIC und BIC. *Minimum information ratio* (MIR) misst die Fähigkeit des Modells, Komponentendichten zu unterscheiden, kann jedoch nicht zwischen einem und mehr Clustern unterscheiden. Das modifizierte MIR (AMIR) ist besser, da MIR die Anzahl an *mixture*-Komponenten unterschätzt. Ziel ist es eine Clusterstruktur des Mischmodells zu finden, welche nicht überlappend und eine klare Komponentenstruktur hat, also das größte Indiz einer Clusterstruktur. Somit wird die Stabilität eines Modells durch die Balance zwischen Modellfit und -komplexität bestimmt. Das *entropy*-Kriterium (Celeux, 1996) beschreibt den *entropy*-Wert der bedingten Wahrscheinlichkeiten unter der Annahme, dass die Beobachtungen einer Population entspringen und misst dabei die Überlappung der Mischmodellkomponenten. Die Mischmodellkomponenten sollen möglichst unterschiedlich sein, dann ist der Wert in der Nähe von eins. Sind sie nicht unterschiedlich, tendiert dieser in die Richtung null. Celeux (1996) vergleicht mit verschiedenen simulierten Daten den *entropy*-Wert NEC, AIC, BIC und ICOMP. NEC zeigt zufriedenstellende Standardabweichungen über die unterschiedlichen Versuche, liegt zwischen AIC und BIC, welche die Anzahl über- bzw. unterschätzen und verhält sich ähnlich wie ICOMP, welcher leicht zur Überschätzung wie AIC tendiert. ICOMP benötigt im Gegensatz zu NEC keine gesonderte Berechnung für den Vergleich zwischen einem und mehr Clustern, jedoch eine zusätzliche Berechnung zum herkömmlichen Schätzoutput. NEC ist einfacher für mehrdimensionale Daten zu rechnen und der EM-*output* kann verwendet werden. Zusätzlich verhält sich NEC besser als ICOMP wenn die Streuung der Cluster größer ist.

4.3.6.15. Fehlerwertbehandlung

Acock (2005) geht detaillierter auf Probleme und Gründe für das Auftauchen von fehlenden Werten ein. *Missing by definition of the subpopulation* bedeutet, dass jene Personen welche fehlende Werte aufweisen per Definition nicht in der interessierten Stichprobe sind. Es gibt vier mögliche Antwortfehlerwerte, nämlich die Antwort wurde verweigert, ‚don't know'-Antwort, gültig übersprungen, oder übersprungen da vom Interviewer vergessen. Gültiges Auslassen von Fragen sollte nicht ersetzt werden. Ist der Interviewer daran schuld, schon. Die ‚don't know'-Kategorie ist am kompliziertesten, da die Person entweder mit der Frage nichts anfangen kann oder die Kategorien nicht vorkommen, die Antwort beispielsweise in der Mitte zweier Kategorien liegt, oder beide Extreme, zufrieden–unzufrieden, gleichermaßen auftauchen. Um fehlende Werte einer Longitudinalstudie als zufällig beschreiben zu können, können über t - und χ^2 -Test Vergleiche zwischen den Personen mit vorhandenen und jenen mit nicht vorhandenen Werten über andere Variablen, welche für beide Gruppen vorhanden sind, gemacht werden. *Missing completely at random* (MCAR) bedeutet zufällige Verteilung der Fehlerwerte über die Datenmatrix, also beispielsweise *missing by design*, wo die Personen nur einen Teil der Fragen gestellt bekommen und die fehlenden Werte nicht von anderen Variablen abhängen. Bei *missing at random* (MAR) hängt die Wahrscheinlichkeit eines fehlenden Wertes nicht vom Wert der Person auf die Frage ab, nachdem die Antworten über andere *mechanism* Variablen kontrolliert wurden. Diese helfen das Fehlen des Wertes zu erklären. *Nonignorable* (NI) *missing values* sind systematisch.

Acock (2005) stellt verschiedene Lösungsansätze dar. *Listwise* oder *case deletion* für MCAR-Daten sind konservativ, da die geringere Stichprobengröße zu höheren Standardabweichungen und niedrigeren Signifikanzgrößen führt und somit Type II Risiko steigt. Dies trifft nicht auf sehr große Stichproben zu. Sind die fehlenden Werte nicht MCAR kann dies zu Verzerrungen führen auch wenn MAR vorliegt. *Pairwise deletion* verwendet alle Daten sofern der Proband ein Paar an Variablen beantwortet hat. Daraus wird die Kovarianzmatrix geschätzt. Diese unterliegt jedoch nicht den Beschränkungen, wie bei Vorliegen aller Daten und somit ist beispielsweise selbst bei MCAR-Daten manchmal die Korrelationsmatrix nicht invertierbar, was jedoch für die Schätzung von Regressions- und Strukturgleichungsmodellen notwendig ist. Auch die Berechnung der Freiheitsgrade ist problematisch. Wird die Stichprobengröße anhand jenes Variablenpaares gewählt, welches die höchste Probandenanzahl aufweist, wird die statistische Kraft überschätzt. Wird das Paar mit der

geringsten Probandenzahl gewählt, wird die statistische Kraft unterschätzt. *Mean substitution* ist unter der Annahme, eine zufällige Beobachtung einer Variablen mit Normalverteilung zu ziehen, angebracht. Sind die fehlenden Werte nicht zufällig, ist sie nicht geeignet, da Personen in der Mitte der Verteilung, verglichen mit am Rande liegenden, meist eine höhere Antwortwahrscheinlichkeit aufweisen. Gibt es viele Fehlerwerte, welche durch den Mittelwert ersetzt werden, werden Standardabweichung/Varianz und Korrelation mit anderen Variablen nach unten verzerrt. Auch der Mittelwertersatz einer Subpopulation, nachdem die Daten anhand anderer Variablen kategorisiert worden sind, senkt die Varianz. Er ist aber ein besserer Schätzer für den Ersatz des fehlenden Wertes und behält mehr von der tatsächlichen Varianz. Bei *indicator/dummy variable adjustment* wird ein binärer Indikator für die Variable erzeugt, welcher angibt, ob der Wert fehlt (1) oder vorhanden ist (0). Dann werden die fehlenden Werte mit dem Mittelwert oder irgendeinem anderen Wert ersetzt. Die Ergebnisse sind die gleichen wie beim listenweisen Fallausschluss, jedoch gibt die Indikatorvariable an, wie weit die mit fehlendem Wert vom Mittelwert der *outcome*-Variablen abweichen. Die zusätzlichen Indikatorvariablen brauchen zwar Freiheitsgrade auf, dieser Verlust führt jedoch zu weit weniger Verlust an Power als die künstlich aufgeblasene Stichprobengröße zu einer Powerzunahme führt. Einführung mehrerer solcher Indikatoren kann zu Multikollinearität führen und bringt andere Schätzungen, verglichen mit dem listenweisen Fallausschluss. Acock (2005) stellt die *single imputation using EM* vor. EM ist ein ML-Ansatz, welcher die Beziehungen zwischen allen Variablen nützt, um fehlende Werte mit einem Grad an Zufallsfehler, welcher die Unsicherheit des Ersatzverfahrens einbringt, zu ersetzen. Die Werte werden iterativ, also die Information der vorherigen Ersetzung nutzend, ersetzt, bis die Kovarianzmatrix sich nicht mehr ändert. *Single imputation* unterschätzt die Standardfehler und überschätzt den Präzisionslevel. Dies liefert mehr Power als die Daten beinhalten. *Multiple imputation*, wobei mehrere separate Datensätze zugerechnet werden, erlaubt *pooling* der Parameterschätzungen um diese zu verbessern und errechnet unverzerrte Standardfehler. *Multiple imputation* errechnet unterschiedliche Lösungen für jede Imputation. Sind diese ähnlich, unterstreicht dies die Imputation. Sind sie unterschiedlich, werden diese Unsicherheiten in die Standardfehler miteinbezogen. Um die Unsicherheiten zu bestimmen, werden ersetzte Datensätze errechnet, damit werden Analysen wie Regression, *propensity score*-Methode oder Markov chain Monte Carlo (MCMC) durchgeführt, die Parameterschätzungen und Standardfehler aufgezeichnet und im Anschluss gepoolt. Poolen bedeutet, dass jede Parameterschätzung der Mittelwert der Anzahl an Replikationen ist, und die Standardabweichung aus der Varianz der unterschiedlichen Lösungen errechnet

wird. Unter MAR liefert *mean imputation* sehr gute Lösungen. FIML ersetzt die fehlenden Werte nicht, nutzt aber die gesamte Information um zu einer ML-Schätzung zu gelangen.

MPlus löst ML-Schätzungen für kontinuierliche, *censored*, binäre, geordnet kategoriale, kategoriale mit mehr als drei Kategorien, Zähldaten oder Kombinationen, gleich ob latent oder nicht. Die Software bietet spezielle Vorgehensweisen für MCAR, MAR oder NI, robuste Standardfehler und *bootstrap*-Standardfehler. Fehlerwert-*patterns* können begutachtet werden um problematisch auftauchende Fehlerwerte zu identifizieren. Fehlen beispielsweise in einer Variablen viele Werte, kann diese eventuell aus der Analyse entfernt werden oder durch eine gleichbedeutende ersetzt werden. FIML und *multiple imputation* wird den traditionellen Verfahren in großen Datensätzen vorgezogen.

4.3.6.16. *Finite growth mixture model* - Kinodatensatz

Ein detailliertes Anwendungsbeispiel latenter Wachstumskurvengruppen anhand der Kinobesucher-Zufriedenheitsstudie wird in einem Journalartikel des Autors vorliegender Dissertation beschrieben (Weismayer, 2010). Es können Gruppen mit unterschiedlichen Wachstumsverläufen aufgefunden werden, woraus eine differenzierte Entwicklung der Zufriedenheitsverläufe resultiert. In den folgenden Tabellen und Diagrammen wird eine kurze Abhandlung des finalen Modells im Artikel beschrieben. Tabelle 73 zeigt die Modellfitevaluierung unterschiedlicher Drei-Klassenlösungen. Die Kreuze in den Spalten I (*intercept*) und S (*slope*) stehen für frei geschätzte Varianzen der beiden Wachstumsfaktoren. Die markierte Zeile beschreibt die best gerankte Lösung. Sowohl der *Vuong-Lo-Mendell-Rubin likelihood ratio test* für zwei (H0) versus drei Klassen, als auch der *parametric bootstrapped likelihood ratio test* für zwei (H0) versus drei Klassen, ist mit einem Wert kleiner 0,0000 hoch signifikant, was die Auswahl von drei verglichen mit zwei Klassen bestätigt.

Eine weitere Modellfitevaluierung bieten die durchschnittlichen latenten Klassenwahrscheinlichkeiten in Tabelle 74, mit den wahrscheinlichsten latenten Klassenzugehörigkeiten in den Zeilen und den latenten Klassen in den Spalten. Hauptdiagonalwerte in der Nähe von eins zeigen ein zufriedenstellendes Ergebnis der Wahrscheinlichkeiten.

	Class 1		Class 2		Class 3		ABIC (sample adjusted BIC)	entropy Wert
	I	S	I	S	I	S		
-	x						51,300.006	0.635
-		x					51,306.153	0.629
-	x	x					51,306.821	0.618
-							51,308.116	0.635
-	x		x				51,383.072	0.663
-	x	x	x				51,415.970	0.665
-		x		x			51,424.786	0.668
-	x			x			51,440.543	0.614
-	x	x		x			51,475.906	0.597
+	x	x	x	x			51,483.965	0.534
+		x		x		x	51,510.540	0.841
+	x		x	x		x	51,531.598	0.920
-	x		x		x		51,539.104	0.647
+	x	x	x	x	x		51,583.424	0.595
+	x	x	x	x		x	51,613.538	0.729

Tabelle 73: Modellfitevaluierung der Drei-Klassenlösung

	Class 1	Class 2	Class 3
Class 1	0.824	0.163	0.014
Class 2	0.195	0.702	0.103
Class 3	0.006	0.102	0.892

Tabelle 74: Latente Klassenzugehörigkeitswahrscheinlichkeiten

Tabelle 75 zeigt die Varianzen innerhalb der drei Klassen. Die Klasse mit dem *random coefficient* weist eine signifikante Korrelation von -0,122 mit *t*-Wert 6,268 auf. In dieser Klasse weisen Personen, welche mit einem hohen Zufriedenheitswert beginnen, eine sinkende Zufriedenheitsentwicklung auf. Die Zufriedenheitswerte jener Personen, welche mit einem niedrigen Zufriedenheitswert beginnen, steigen über die Zeit hinweg.

Diagramm 126 zeigt die Drei-Klassenlösung. In dieser sind *intercept*- und *slope*-Werte in zwei der drei Klassen auf null fixiert, da diese für *random effects* nicht genügend Streuung aufweisen.

	Class 1		Class 2		Class 3	
	I	S	I	S	I	S
Varianz	0.000	0.000	0.000	0.000	0.952	0.028
Standardabweichung	0.000	0.000	0.000	0.000	0.135	0.003
t-Wert	-	-	-	-	7.040	8.009

Tabelle 75: Latente Klassenvarianzen

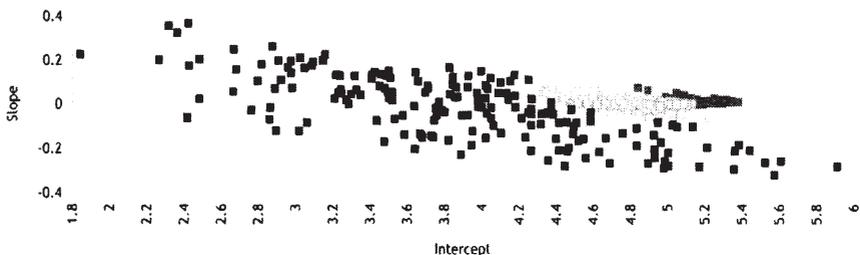


Diagramm 126: Drei-Klassenmodell latenter Wachstumskurven

4.3.7. *Markov Chain Modelling: von der latent class analysis zum mixed latent Markov chain model*

Collins & Wugalter (1992) testen Modelle, welche die klassischen Markov-Modelle um komplexere latente Variable mit mehreren Indikatoren erweitern. Der Vorteil liegt in dynamisch latenten Variablen, anstatt statisch latenter, welche Entwicklungsphasen in *stage-sequential*-Prozessen besser beschreiben. Mehrindikatormodelle mit *stage-sequential* latenten Variablen in der latenten Klassentheorie nennen sie *latent transition analysis* (LTA). Die bis dahin vorhandenen Markov-Modelle beinhalten großteils nur einen manifesten Indikator. Ziel ist die stufenweise Übergangswahrscheinlichkeit vorherzusagen. Die Übergangswahrscheinlichkeitsmatrix, *transition probability matrix*, zeigt die Wahrscheinlichkeiten der Zugehörigkeit zu Stufe zwei am Ende eines Intervalls, gegeben die Zugehörigkeit zu Stufe eins am Beginn des Intervalls. Markov-Modelle bilden die Daten aber nur adäquat ab, wenn Antwortunsicherheit erlaubt wird. Deshalb werden sie mit latenten Klassen kombiniert und folglich steigern mehrere Items die Reliabilität und Validität. In genanntem Artikel wird ein *time homogeneous first-order Markov*-Modell gerechnet. Zu jedem Beobachtungszeitpunkt existiert für jedes Individuum eine latente Klassenzugehörigkeit im Sinne eines latenten Zustandes, *latent status*. In der LTA sind die Zeilen und Spalten der Übergangswahrscheinlichkeitsmatrix selbst latente Zustände und keine *stages*. So können die latenten Zustände kumulative oder *stage-skipping*-Entwicklungen annehmen. In ersterem Fall werden neue Fähigkeiten erworben und die zuvor vorhandenen bleiben erhalten bzw. gehen Fähigkeiten verloren,

bleiben diese verloren. In letzterem nicht kumulativem Fall verlieren die Personen die Charakteristiken der Vorperiode. Collins & Wugalter (1992) stellen eine kumulativ monotone Entwicklung und eine kumulative Entwicklung, welche Entwicklungsrückschritte erlaubt, vor. Geschätzt werden der Anteil der Gruppe, der Anteil der Personen auf den latenten Zuständen zum Zeitpunkt t gegeben die Gruppe, und die Wahrscheinlichkeit der Zugehörigkeit zu einem latenten Zustand zum Zeitpunkt $t+1$ gegeben die Zugehörigkeit zu einem latenten Zustand zum Zeitpunkt t und Zugehörigkeit zu einer Gruppe. LTA bildet die Items als latenten Zustand ab und gibt die Messgenauigkeit aus.

Geschätzt wird das Modell mit dem EM-Algorithmus in zwei Schritten. Der *E-step* berechnet wiederum die erwartete suffiziente Statistik, die Antwortmusteranteile, basierend auf den aktuellen Parameterschätzungen. Der *M-step* errechnet durch Maximierung der *likelihood*-Funktion neue Parameterschätzungen auf Basis der aktuell geschätzten suffizienten Statistik. Collins & Wugalter (1992) betonen die Brauchbarkeit des EM-Algorithmus, wenn unterschiedliche Startwerte verwendet werden, auch für latente Klassenmodelle und die Nichtnotwendigkeit extrem großer Stichproben für latente Markov-Modelle mit mehreren Indikatoren. Die Annahme schlechterer Parameterschätzungen durch Hinzunahme von Items, da die einzelnen Häufigkeiten der Antwortpatterns kleiner werden, wird nicht bestätigt. Es zeigt sich, dass zusätzliche Items die Standardabweichungen senken und der geschätzte Mittelwert näher beim beobachteten liegt. Bei wenig Items werden größere Stichproben empfohlen.

Durch Fixierung eines Parameters auf einen Wert oder Gleichsetzung mehrerer Parameter wird das Modell schlanker und leichter spezifizierbar. Werden beispielsweise Entwicklungsrückschritte verboten, wird die untere Dreiecksmatrix der Übergangswahrscheinlichkeitsmatrix auf null fixiert. Empfohlen wird auch die Charakterisierung der latenten Zustände um die Konstrukte über die Zeit vergleichbar zu machen. Auch ein Rate- oder Vergessensparameter kann über Items und/oder Beobachtungen gleich gesetzt werden. Modellfit wird über die *LR-statistic*, welche den Unterschied zwischen der Anzahl an Antwortmustern und Anzahl an geschätzten Parametern minus eins angibt, oder über AIC bestimmt. Durch Simulation wird die Stabilität unterschiedlicher Modellansätze getestet, nämlich eine kumulative Entwicklung ohne Entwicklungsrückschritte und eine nicht-kumulative mit möglichen Entwicklungsrückschritten. In ersterem Fall stellt die obere Dreiecksmatrix die Übergangswahrscheinlichkeitsmatrix dar und die restlichen Parameter werden fixiert. In letzterem Fall werden alle Parameter frei geschätzt. Zusätzlich werden Messstärke, Itemanzahl und Stichprobengröße variiert. Beschränkungen werden

beispielsweise hinsichtlich der Übergangswahrscheinlichkeiten über die Zeit eingeführt. Auch Restriktionen in Bezug auf ein Item, welche die Unsicherheit der Messung des latenten Zustandes über alle latenten Zustände gleichsetzen, werden eingeführt. Drei Modelle werden empirisch getestet. In einem sind Lernfortschritte aber keine Rückschritte erlaubt. Dies führt zur Schätzung der oberen Dreiecksmatrix, auch *longitudinal Guttman simplex* (LGS) genannt. In einem weiteren werden alle Matrixparameter geschätzt und im letzten wird nichts erlaubt und die Einheitsmatrix fixiert. Es existieren fünf latente Zustände und somit werden zu Beginn für diese fünf latenten Zustände Personenanteile geschätzt. Für jede Beobachtungskombination gibt es eine Wahrscheinlichkeit (latente Zustände x Items x Antwortkategorien x Beobachtungen), wobei eine Kategorie davon nicht geschätzt werden muss. Weiters werden gleiche Wahrscheinlichkeiten über die Zeit angenommen. Da die Veränderung sehr langsam vor sich geht, wird ein Modell erstellt, bei welchem alle Elemente abgesehen von der Hauptdiagonale und dessen darauffolgendem höheren latenten Zustand fixiert werden. Verglichen wird mit LR-Differenztest und AIC. Aus der geschätzten Hauptdiagonale kann keine Veränderungswahrscheinlichkeit abgelesen werden. Collins & Wugalter (1992) vergleichen LGS mit LTA. Ersterer beschreibt eine dynamisch latente Variable, die einer kumulativen einheitlichen und irreversiblen Entwicklung unterliegt, axiomatisch. Das heißt, die Skalen werden daran evaluiert, wie oft die manifesten Daten den Modellaxiomen folgen oder dies nicht tun. Es gibt es keinen Abgleich oder Schätzung der Fehler. Letztere schätzt eine Fehlerstruktur und eine latente Übergangswahrscheinlichkeitsmatrix. Ein Entwicklungsrückschritt widerspricht den LGS-Annahmen, jedoch nicht denen der LTA. Nachteile der LTA sind große Kontingenztabellen bei vielen Beobachtungen oder die oft verletzte Annahme der χ^2 -Verteilung der *likelihood statistic* selbst für genestete Vergleiche. Vorteile liegen in der Entwicklungstestung beim Gebrauch mehrerer Indikatoren und einhergehender Senkung der Standardfehler und in der Verwendung als Prozedur zur Instrumententwicklung durch Ausschluss schlechter Items. Die Konstruktvalidität der Messung *stage-sequential* dynamisch latenter Variablen wird verbessert.

4.3.7.1. Anwendungsbeispiele und Spezifizierungen

Collins et al. (1994) stellen einen Kreuzvalidierungsansatz für Kovarianzstrukturmodelle vor. Das Problem liegt in der Auffindung der Approximation/en, welche dem/den nicht bekannten wahren Modell/en am ähnlichsten ist/sind. Eine Möglichkeit ist die zufällige Teilung des Datensatzes in zwei Hälften. Bevorzugen die Fitindizes der getesteten Modelle beider Schätzungen das gleiche Modell, kann dies als beste Approximation angesehen werden. Der Grund für die

Ausweichnotwendigkeit liegt in der Verletzung der χ^2 -verteilten LR-*statistic* oder der Person- χ^2 -Verteilung durch die auftretenden leeren und schwach besetzten Zellen in der Kontingenztabelle. Monte Carlo-Studien zeigen, dass die Verteilung der Differenz der LR-*statistic* genesteter Modelle nicht sehr robust ist. Werden externe Gruppen miteinander verglichen, z.B. Geschlecht, stellt sich die Frage, ob diese aufgrund der Gruppierung oder aufgrund von Stichprobenfluktuation auftreten. Die genesteten Modelle, eines mit frei geschätzten und eines mit gleichgesetzten Parametern zwischen den Gruppen, können über die LR-*statistic* ebenfalls nur mit den gleichen innewohnenden Problemen getestet werden.

Collins et al. (1994) führen eine LTA durch. Sie entspricht dem latenten Klassenansatz für Modelle von *stage-sequential* dynamisch latenten Variablen. Die latente Klassenzugehörigkeit eines Individuums zu einem gewissen Zeitpunkt entspricht dem latenten Zustand. LTA misst Veränderungen zwischen latenten Statuszugehörigkeiten über die Zeit. Natürlich entspricht nicht jedes Antwortpattern einer eigenen Klasse. Somit gibt es keine eins-zu-eins Abbildung der Antwortpatterns und der latenten Zustände, sondern eine Wahrscheinlichkeit dafür. Angenommen jemand ist zu jedem Zeitpunkt unzufrieden, so kann die Wahrscheinlichkeit der Zufriedenheit trotzdem höher als null liegen. Es existieren meist auch Antwortpatterns, welche sehr schlecht zu den latenten Zuständen passen. LTA bezieht Messfehler durch Wahrscheinlichkeiten der latenten Klasse und latenten Zustandszugehörigkeit mit ein.

Ein Konvergenzkriterium für die Schätzung liegt in der mittleren absoluten Abweichung der Parameterschätzungen aufeinanderfolgender Iterationen. *Goodness-of-fit* Bestimmung erfolgt über die LR-*statistic*. Um lokale Maxima zu vermeiden, werden unterschiedliche Startwerte gewählt. Werden unterschiedliche Lösungen gefunden, werden diese für die vorgeschlagene Kreuzvalidierung verwendet. Für beide Hälften wird das Modell separat geschätzt und die Daten des anderen Modells zum Vergleich herangezogen. So dient jede Datenhälfte sowohl zur Kalibrierung, als auch zur Kreuzvalidierung. Um Geschlechterunterschiede beurteilen zu können, muss für beide Geschlechter das gleiche Modell mit der vorgeschlagenen Sequenz der latenten Zustände bevorzugt werden, sonst führen unterschiedliche Prozesswege der Geschlechter zu Verzerrungen. Sind diese gleich, so werden *multiple group*-Modelle mit unterschiedlichen Einschränkungen verglichen, wobei das Geschlecht als exogene Variable dient. Im ersten Modell wurden keine Beschränkungen vorgenommen. Somit ist in diesem Modell die Anzahl und Sequenz der latenten Zustände die gleiche, aber die Bedeutung der latenten Zustände unterscheidet

sich zwischen den Geschlechtern. Im zweiten Modell werden die latenten Zustände als gleich über die Geschlechter fixiert, um die Bedeutung über die Geschlechter ebenfalls gleich zu halten. Jedoch werden unterschiedliche Startwerte und Veränderungen erlaubt. Im dritten Modell werden die Veränderungsraten gleichgehalten und Variation nur über die Startwerte erlaubt. Das vierte Modell ist das vice versa-Modell vom dritten. Im fünften Modell wird alles gleich gehalten und es werden keine Geschlechterunterschiede erlaubt.

Einige Beschränkungen werden eingeführt und verkleinern die Komplexität der *full transition probability*-Matrix. So werden Rückentwicklungen durch Nullsetzen der unteren Dreiecksmatrix unterdrückt. Sind diese ohnehin nicht logisch, sollten sich niedrige Werte ergeben. Schließlich werden die Startgrößen und Veränderungswahrscheinlichkeiten der beiden Datenhälften und auch zwischen den Geschlechtern verglichen. Unterscheiden sich die Schätzungen zwischen den Datenhälften nicht, wird von einer guten Modellqualität ausgegangen. Weiters dient die Kreuzvalidierung zwischen den Wahrscheinlichkeiten einen Wert für ein Item in einer Gruppe zu beobachten zur Bestimmung der Interpretationsähnlichkeiten der latenten Zustände zwischen den Teilstichproben. Collins et al. (1994) merken an, dass bei der Untersuchung von Gruppendifferenzen der Untersuchungsgegenstand über die Gruppen als gleich betrachtet werden muss und die Kreuzvalidierung eine geeignete Methode dafür ist. In ihrem Beispiel würden unterschiedliche Alkoholismusabhängigkeitsprozesse zwischen Geschlechtern, welche als gleich angenommen werden, jedoch möglicherweise fälschlicherweise zu falschen Preventionsprogrammen führen. Abschließend ist der wichtigste Punkt bei Verwendung der LTA die sorgfältige Modellauswahl. Es können einfache geordnete Entwicklungen, aber auch komplexe nicht geordnete existieren.

Meiser & Ohrt (1996) beschreiben aus der Familie der *finite mixture distribution*-Modelle, welche unterschiedliche Entwicklungsprozesse verschiedener latenter Subpopulationen erlauben, eine Spezifikation im Rahmen der *mixed latent Markov chains* (MLM) mit mehreren Indikatoren pro Zeitpunkt. Diese ermöglicht gleichzeitig *random fluctuations*, also zufällige Fluktuationen zwischen Zuständen einer latenten Subpopulation, zusammen mit stabilen oder kohärenten Entwicklungstrajektorien anderer Subpopulationen ein einem Modell. *Finite mixture distribution*-Modelle unterteilen die Population in latente Subpopulationen mit eigenen Entwicklungspatterns und gleicher Verteilungsart. Alle Komponentenverteilungen gehören der gleichen parametrischen Familie an und jede Subpopulation besitzt ihre eigenen spezifischen Parameter. Für diskrete

longitudinale Daten werden nur *mixtures* von parametrischen multinomialen Verteilungen betrachtet, wie beispielsweise in der *latent class analysis* (LCA).

Meiser & Ohrt (1996) stellen *finite mixture multinomial*-Modelle vor, welche unterschiedliche Parametrisierungen verschiedener Subpopulationen zulassen, nämlich *latent Markov* und *marginal LCA*-Modelle, zusammen auch *mixed latent partial Markov-chain models* (MLPM) genannt. *Random fluctuation* oder *transition by chance* bedeutet, dass die Klassenzugehörigkeit zum Zeitpunkt t nicht vom Zeitpunkt $t-1$ abhängt und auch nicht die folgende beeinflusst. Folglich sind die latenten Zustände stochastisch unabhängig und die Wahrscheinlichkeit einer bestimmten latenten Klasse anzugehören, hängt zum Zeitpunkt t lediglich von der marginalen Verteilung der latenten Klassen zum selbigen Zeitpunkt ab. Systematische Veränderungen sind Entwicklungsprozesse mit stochastisch aufeinander bezogenen latenten Zuständen über die Zeit oder bestimmte Übergänge zwischen zwei latenten Zuständen, wie beispielsweise beim kumulativen Wachstum mit geordneten latenten Klassen ohne Rückentwicklungsmöglichkeit.

In *latent Markov chain* (LM)-Modellen gibt es für jeden Beobachtungszeitpunkt ein unidimensionales Messmodell im Sinne einer LCA. Das longitudinale LM bezieht die bedingte stochastische Unabhängigkeit aller beobachteten Indikatoren, gegeben den Vektor der latenten Klassen, mit ein. Weiters wird der Wahrscheinlichkeit des latenten Vektors, ein autoregressives Modell erster Ordnung unterstellt. Das Messmodell besteht aus Antwortwahrscheinlichkeiten, welche das Auftreten einer Kategorie eines bestimmten Indikators zu einem bestimmten Zeitpunkt gegeben der latenten Klasse beschreiben, und aus der *a priori* bestimmten Anzahl an Klassen. Das Strukturmodell besteht aus Übergangswahrscheinlichkeiten, also dem Anteil der latenten Klasse zu einem bestimmten Zeitpunkt gegeben den Wert der latenten Variable zum vorangehenden Zeitpunkt, und aus den *initial class proportions*, also der anfänglichen Verteilung der latenten Variable zum ersten Zeitpunkt.

MLM sind *finite mixture*-Modelle mit LM-Komponenten. Eine diskrete latente *mixing*-Variable teilt die Population in Subpopulationen, genannt Ketten. Ein Einketten-MLM ist beispielsweise ein LM. Jede dieser Ketten bildet ein eigenes LM, gegeben die diskrete latente *mixing*-Variable. Somit existiert eine *a priori* bestimmte Anzahl an Ketten mit Kettenanteilen. Beschränkungen innerhalb von Ketten, also Übergangswahrscheinlichkeiten welche latente Stabilität definieren, und über Ketten, also gleiche Messmodelle, damit nur die Strukturmodelle variieren, oder zu den Antwortwahrscheinlichkeiten über die Zeit, sind wiederum

möglich. Im *mixed latent partial Markov-chain model* können ebenfalls für einzelne Ketten stochastische Unabhängigkeiten benachbarter latenter Zustände definiert werden. Die Übergangparameter beschreiben dadurch die marginale Verteilung der latenten Variablen zu einem bestimmten Zeitpunkt. Werden in einer Kette alle latenten Variablen der unterschiedlichen Zeitpunkte so verbunden, stellt dies eine *chain of transitions by chance* dar. Unabhängigkeit zwischen den *marginals* der Messbeobachtungen, also zwischen den Randverteilungen, wobei jede *marginal* eine LCA, gegeben die Aufteilung in Markov-Ketten ist, führt zu Identifizierungsproblemen. Durch zusätzliche Beschränkungen innerhalb oder zwischen Ketten wird jede marginale LCA eigens identifiziert. Meiser & Ohrt (1996) definieren in ihrem empirischen Beispiel drei Subpopulationen. Die erste Kette ist ein *latent stayer*-Modell mit latenter Stabilität, die zweite eine mit anhaltender Behandlung unter Ausschluss unlogischer Rückentwicklung, also keine Behandlung zum Zeitpunkt zwei wenn es zum Zeitpunkt eins bereits eine Behandlung gegeben hat, und die dritte erlaubt Behandlungen zu jedem Zeitpunkt, ohne diese jedoch für spätere Zeitpunkte vorhersagen zu können. Die dritte geht also von Unabhängigkeit zwischen Zeitpunkten aus. Das Ergebnis liefert Kettenanteile, Startanteile, Übergangswahrscheinlichkeiten für die einzelnen Ketten, wobei letztere in der dritten Gruppe gleich sind, und die bedingten Antwortwahrscheinlichkeiten über alle Ketten. Letzte stehen für die Reliabilität. Über LRT-*statistic*, AIC und BIC werden ein Drei-Kettenmodell, ein Zwei-Kettenmodell ohne *random fluctuation*-Kette und ein Ein-Kettenmodell ohne Beschränkungen, verglichen.

4.3.7.2. Die Modellierungsumgebung - *General Growth Mixture Modelling*

Muthén und Muthén (2000) beschreiben theoretisch die unterschiedlichen Modelle. Die LCA hat zum Ziel, durch Einführung einer kategorialen latenten Variable, die kleinst-mögliche Gruppe zu finden, welche die beobachteten Daten gerade noch beschreibt. Die geschätzten Klassenwahrscheinlichkeiten werden *posterior probabilities* genannt.

Die LTA untersucht Klassenzugehörigkeitsveränderungen über die Zeit. Es existiert eine latente Klassenvariable zu jedem Zeitpunkt und die Übergangswahrscheinlichkeit zwischen Klassen zu zwei Zeitpunkten wird berechnet.

Latent class growth analysis (LCGA) dient zur Aufdeckung latenter Klassen mit unterschiedlichen Wachstumsverläufen einer einzigen Variablen die wiederholt gemessen wurde. Unterschiedliche Wachstumsverläufe und Klassenwahrscheinlichkeiten werden geschätzt und die *posterior*-Klassenzugehörigkeitswahrscheinlichkeiten für alle Individuen berechnet.

Growth mixture modeling (GMM) inkludiert *within-class*-Variation durch *random coefficients*, also Zufallskoeffizienten, welche kontinuierlich latente Variablen oder Wachstumsfaktoren darstellen und die Individuenvariation messen. Wachstumsfaktorvarianzen werden für jede Gruppe geschätzt.

General growth mixture modeling (GGMM) inkludiert sämtliche eben genannte Analysen wie in Diagramm 127 dargestellt, beispielsweise auch *distal outcomes*, welche durch die Klassenzugehörigkeit vorhergesagt werden. Kreis A enthält die Faktorenanalyse und *conventional growth modeling* mit kontinuierlichen Daten, Kreis B enthält LCA, LTA und LCGA unter Verwendung kategorialer Daten, Kreis C enthält GMM, welches die Möglichkeit einer latenten kategorialen Klassenvariable inkludiert und Kreis D beschreibt die Verbindung sämtlicher Modelle. Kreis D erweitert die traditionellen latenten Variablenmodelle, zusätzlich zu den kategorialen latenten, um kontinuierliche latente Variablen. Muthén und Muthén (2000) zeigen die Beziehung zwischen LCA und EFA auf. LCA verwendet kategoriale Daten, EFA verwendet kontinuierliche Daten. Individuen unterschiedlicher Klassen weisen unterschiedlich hohe Werte in den errechneten Faktoren in der EFA auf. Dennoch sind die Ladungen der EFA nicht gleich den Wahrscheinlichkeiten der LCA, welche kategoriale Daten verwendet. Beim Vergleich zwischen LCGA und GMM spielen zusätzliche Kriterien eine Rolle, wie beispielsweise das zeitliche Zusammenhängen über Autokorrelationen oder die Festlegung gleichbleibender oder wechselnder Übergänge.

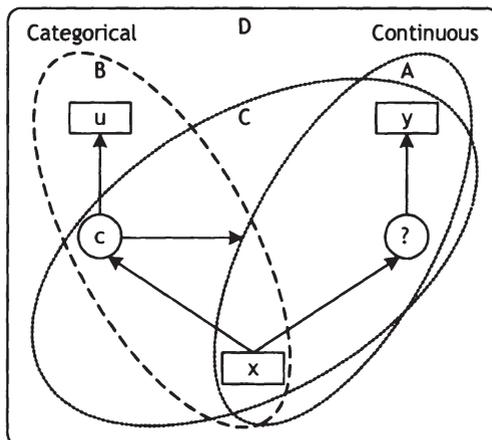


Diagramm 127: General modeling framework (Muthén und Muthén, 2000)

BIC sucht die Balance zwischen Maximierung der *likelihood* und Modellsparbarkeit und dient zur Modellbeurteilung. Weiters wird die Qualität der Klassifikation durch Begutachtung der durchschnittlichen *posterior probabilities* für die Klassen bestimmt. Die Werte der Hauptdiagonalen müssen klar höher sein, als die in den anderen Feldern. Auch die praktische Brauchbarkeit der Klassen, also die Ähnlichkeiten von Wachstumskurven, die Individuenzahl pro Klasse, da die Generalisierung weniger Individuen fraglich ist, die Anzahl der geschätzten Parameter, da eine geringere Parameteranzahl sich in einer neuen Stichprobe eher bewahrheitet und unterschiedliche Vorhersagen bei Verwendung unterschiedlicher Klassenanzahl, spielen eine Rolle.

Muthén und Muthén (2000) stellen zwei GGMM vor. Im ersten werden Kovariate miteinbezogen und *odds ratios* zum Vergleich mit einer Gruppe errechnet. Die multinomial logistischen Regressionskoeffizienten einer Gruppe werden also auf null standardisiert und in Bezug auf die anderen Klassen über die *log odds* verglichen. Im zweiten Beispiel wird eine erweiterte Form der LTA, welche latente Klassenvariablen auf Basis mehrerer Indikatoren zu mehreren Zeitpunkten nützen, verwendet, nämlich durch die Verbindung zweier latenter Klassenvariablen, welche beide unterschiedliche Konstrukte beschreiben. Hier werden die bedingten Wahrscheinlichkeiten, die Übergangswahrscheinlichkeiten der Klassenzugehörigkeit der stark Alkoholabhängigen, als Funktion der Klassenzugehörigkeit zum antisozialen Verhalten geschätzt. Die marginalen Wahrscheinlichkeiten der beiden Klassifikationen und die bedingten Wahrscheinlichkeiten, also die *odds* über die beiden Klassen, beschreiben die Beziehung zwischen beiden Klassifikationen.

4.3.7.3. Modellierungsmöglichkeiten

Vermunt (2007) beschreibt unterschiedliche Formen von Wachstumsmodellen. Dabei wird die bivariate Abhängigkeit zwischen Beobachtung und beieinander liegenden Beobachtungszeitpunkten von *transitional models*, also Übergangsmodellen, erwähnt. Wachstumsmodelle gehen davon aus, dass zeitspezifische Antworten eine Funktion der Zeit sind, somit ein Mitglied der GLM-Familie. Beobachtungen des gleichen Individuums zu verschiedenen Zeitpunkten sind nicht unabhängig. Deshalb sind manche Parameter gruppenspezifisch. Zwecks Sparsamkeit wird den Zufallskoeffizienten multivariate Normalverteilung unterstellt. Im *mixed modeling*-Ansatz, *mixed-effects* GLM, werden die *fixed* und *random effects* getrennt. Geschätzt wird über ML, wobei es nützlich ist, das *mixed-effects* Modell als faktoranalytisches Modell mit nicht korrelierten latenten Variablen als Zufallseffekte mit Varianz eins zu parametrisieren. Er kann für kontinuierliche, binäre oder Zähldaten

verwendet werden. Auch Lösungen für *ordinal mixed-effects* GLM und nominale werden vorgestellt. Es wird die Erweiterung über *time-constant* (*between-subject*) und *time-varying* (*within-subject*) Prädiktoren, also zeitkonstante und zeitvariante, vorgestellt. Nachteile liegen im Rechenaufwand bei kategorialen Antwortvariablen und die Annahme multivariat normalverteilter Zufalls-koeffizienten. Wird die Verteilungsannahme verletzt, können die Ergebnisse verzerrt sein.

Vermunt (2007) beschreibt *latent-class-based and hybrid growth*-Modelle über einen *latent-class-based nonparametric random coefficients*-Ansatz, auch *latent class* (LC) oder *mixture regression analysis* genannt. Der erwartete Wert einer Person hängt von der latenten Klasse ab und ist eine Funktion der Zeit, wobei die Parameter über die Klassen variieren. Es gibt klassenspezifische Wachstumskurven, welche gleich der Anzahl der Klassen bzw. der *mixture*-Komponenten sind. Um zeitkonstante Kovariate einzubringen gibt es zwei Möglichkeiten, nämlich im Regressionsmodell der *response*-Variablen oder im dem der latenten Klassen. Vorteile des Ansatzes von Vermunt (2007) liegen im geringeren Rechenaufwand bei kategorialen Daten und in der Interpretierbarkeit mehrerer Trajektorien. Schwachpunkte liegen in notwendigen Entscheidungen, welche Parameter über die Klassen variieren dürfen, Anzahl der Klassen und funktionale Form der klassenspezifischen Zeitfunktion.

Zuerst werden alle Klassen mit gleicher Form geschätzt und dann über BIC die Anzahl der Klassen bestimmt. Schließlich wird durch Gleichsetzung von Parametern über Klassen oder Fixierung von Koeffizienten auf null das Modell sparsamer gestaltet. Das Risiko von lokalen Maxima in nicht-parametrischen *random-effects*-Modellen ist höher, verglichen mit parametrischen Modellen. Oft wird eine zu große Anzahl latenter Klassen benötigt um die Variation aufzufangen, welche aber meist nur die des *intercepts* betrifft. Einen Ausweg liefert die *hybrid methodology*, wobei diskrete mit kontinuierlicher unbeobachteter Heterogenität kombiniert wird. Das LC-Regressionsmodell wird um einen *random intercept* erweitert. Dies weicht die Basisannahme von LC-Modellen, dass die latenten Klassen homogen bezüglich aller Modellparameter sein müssen, auf. Zusätzlich wird *within-class*-Heterogenität des *intercepts* erlaubt. So gibt es zwei Erweiterungen, nämlich die *random change*-Parameter-Inkludierung zusätzlich zum *random intercept* und die Erlaubnis der variierenden Varianz über die latenten Klassen. Ein derartiges Modell geht von *within-class*-Heterogenität aus und auch davon, dass der Betrag der *within-class*-Heterogenität über die Klassen variiert.

Vermunt (2007) liefert ein Beispiel zur Schizophrenie. Verschiedene Modelle werden anhand unterschiedlicher Modelleigenschaften variiert. So werden beispielsweise die Arten wie die Modelle unbeobachtete Heterogenität auffangen, wie sie den Aspekt der *treatment*-Gruppe vs. Kontrollgruppe in das Modell aufnehmen (also *treatment* beeinflusst den *response/outcome* vs. die Klassenzugehörigkeit), oder die Anzahl an Klassen und Wachstumsverläufen in den Klassen, von Modell zu Modell verändert. Die Modelle werden über den *log-likelihood*-Wert, BIC und *Wald*-Test für die *treatment*-Gruppe verglichen. Letzterer testet die zeitliche *treatment*-Interaktion bzw. den Effekt auf die Klassenzugehörigkeit. Der *treatment*-Effekt ist nicht signifikant, wenn die nicht beobachtete Heterogenität nicht miteinbezogen wird. Ein visueller Vergleich zwischen beobachteten und geschätzten Wahrscheinlichkeiten zeigt eine geglättete Entwicklung mit größeren Unterschieden. Schließlich werden drei unterschiedliche Markov Modelle geschätzt und verglichen.

Wachstumsmodelle beschreiben individuelle Unterschiede in der Wahrscheinlichkeit zu einem Zeitpunkt einen gewissen Level zu besitzen und dies in Abhängigkeit der Zeit oder anderen Prädiktoren. Markov-Modelle beschreiben aggregierte Übergangswahrscheinlichkeiten, also die gesamte Wahrscheinlichkeit, einen bestimmten Zustand gegeben den vorangegangenen Zeitpunkt zu besitzen. Für die erwähnten Modelle entspricht dies dem Wahrscheinlichkeitsunterschied der Kontroll- vs. *treatment*-Gruppe, einen Übergang zwischen Schizophrenie und Gesundheit zu erleben. *Mixed* Markov und *latent class* Markov sind zwei LC-basierte Varianten des *standard* Markov-Modells. In ersterem gehört das Individuum einer Klasse an, wobei sich die Klassen anhand ihrer Übergangswahrscheinlichkeiten unterscheiden. In letzterem wird die *stage*-Variable mit einem Fehler gemessen. Dadurch werden Übergänge zwischen zwei latenten Zuständen modelliert, welche zu den beobachteten Zuständen mit einem Wahrscheinlichkeitsmechanismus verbunden werden. Vermunt (2007) modelliert ein *standard stationary first-order* Markov-Modell, ein *mixed* Markov-Modell mit zwei latenten Klassen und ein LC-Markov-Modell mit zwei latenten Zuständen pro Zeitpunkt. Die Fits sind genauso gut wie der beste der Wachstumsmodelle. Ähnliche *log-likelihood* Werte der drei Modelle weisen auf ein zufriedenstellendes *standard* Markov-Modell hin. Somit gibt es keine Hinweise auf unbeobachtete Heterogenität in den Übergangswahrscheinlichkeiten, wie sie im *mixed* Markov-Modell angenommen wird. Ebenfalls besteht keine Notwendigkeit für Messfehler im zeitspezifischen Zustand wie im LC Markov-Modell. Die geschätzten Übergangswahrscheinlichkeiten des *stationary first order* Markov-Modells zeigen eine höhere Übergangswahrscheinlichkeit der *treatment*-Gruppe hinaus

aus dem Schizophreniezustand und eine niedrigere Wahrscheinlichkeit hinein in den Schizophreniezustand.

Vermunt (2007) erwähnt weiters die Möglichkeit mehrerer Hierarchien, nämlich multiple Antworten genestet in Zeitpunkten und Zeitpunkte genestet in Fällen, und Zeitpunkte genestet in Fällen und Fälle genestet in Gruppen. Es wird angemerkt, dass abhängig davon, ob ein *standard* oder LC-basiertes Wachstumsmodell verwendet wird, für diesen Teil in erstem Fall ein *mixed-effect* multinomial logistisches Regressionsmodell und im zweiten ein LC multinomial logistisches Regressionsmodell entsteht. Der *multilevel* LC-Ansatz für *multiple response growth modeling* ist nützlich, wenn die multiplen Antworten Indikatoren einer latenten Variable darstellen.

4.3.7.4. Klassische Modellhierarchie vom *mixture latent Markov model* ausgehend

Vermunt et al. (2008) listen unterschiedliche LC-Modelle für longitudinale Daten auf. Ihr Modell ist für Datensätze geeignet, in welchen die Zeit eine diskrete Variable darstellt. Zur Schätzung wird ML durch eine Abwandlung des EM-Algorithmus, nämlich eine *forward-backward* oder Baum-Welch-Lösung präsentiert, welcher die Behandlung von fehlenden Werten unter MAR-Anahme erlaubt, jedoch nicht in der Kovariaten.

Das *mixture latent Markov*-Modell ist ein LC-Modell mit *time-constant* latenten Variablen, welche nicht beobachtete Heterogenität eines Veränderungsprozesses messen, und *time-varying discrete* latenten Variablen, welche Messfehler der beobachteten Antworten mit einbeziehen. Verbindungen zwischen den latenten Zuständen aufeinanderfolgender Zeitpunkte bildet die *Markovian transition* oder die autoregressive Korrelationsstruktur erster Ordnung.

Im *standard* LC-Modell werden Beobachtungen als unabhängig über die Zeitpunkte angenommen. Somit ist das Modell ohne latente Klassen und Zustände lediglich für *cross-sectional* Daten geeignet.

Im *mixture growth model* (*mixture* oder *latent class growth model*) gibt es nur einen Indikator, wobei die Kategorien des Indikators den latenten Zuständen entsprechen. Durch 1- bzw. 0-Zuteilung gilt die Annahme, dass ohne Fehler gemessen wird.

Im *mixture latent growth model* (*non-parametric random-effects model*) gibt es keine Veränderungen oder Markovstrukturen und somit lediglich die *time-*

constant latente Variable. Die dynamische abhängige Variable wird zur (diskreten) latenten Variable, welche durch mehrere Indikatoren gemessen wird.

Im Markov *model* wird eine perfekte Messung wie im *mixture Markov model* angenommen. Es gibt keine unbeobachtete Heterogenität wie im *latent Markov model*. Somit entspricht dieses Modell nicht mehr einem latenten Klassenmodell, sondern einem Modell mit einer Antwortvariablen und einer Markov-Struktur.

Im *latent Markov model* (*latent Markov, latent transition, hidden Markov model*) gibt es keine *time-constant* latente Variable. Es gilt die Annahme, dass es keine unbeobachtete Heterogenität gibt und alle Subjekte derselben latenten Klasse angehören. Geschätzt werden der Zustand zu Beginn und die Veränderungswahrscheinlichkeiten. Gibt es nur eine beobachtete Variable, so entspricht die Anzahl der Kategorien der beobachteten Variablen, den latenten Zuständen. Aus Identifizierungszwecken können zeitlich gleichbleibende Veränderungswahrscheinlichkeiten oder Fehlklassifizierungswahrscheinlichkeiten angenommen werden. Gibt es mehrere beobachtete Variablen ist das Modell eine Erweiterung des *standard LC-Modells*. Die zeitspezifischen latenten Zustände werden als latente Cluster oder Typen angesehen, welche sich in ihrem Antwortverhalten unterscheiden. Die Markovianische Veränderungsstruktur beschreibt die Veränderungen zwischen benachbarten Beobachtungen.

Im *mixture Markov model* gibt es nur eine Antwortvariable, wobei angenommen wird, dass diese ohne Messfehler beobachtet wird. Die Antwortkategorien entsprechen wiederum den latenten Zuständen mit Zuteilung 0 bzw. 1. Im *mover-stayer* Modell wird die Übergangswahrscheinlichkeit einer Gruppe, der *stayer*-Gruppe, konstant auf null gehalten.

Das *mixed Markov latent class model* beinhaltet weder *time-constant* noch *time-varying* Kovariate.

4.3.7.5. Modellspezifikationen des umfassenden *mixture latent Markov model*

Zuvor genannte Modelle, stellen Restriktionen des *mixture latent Markov-Modells* dar, welches fünf unterschiedliche Variablen enthält, *response variables, time-constant explanatory variables, time-varying explanatory variables, time-constant discrete latent variables* und *time-varying discrete latent variables*. Das Modell beinhaltet eine zeitkonstante latente Variable und $t+1$ zeitvariierende latente Variablen. Die Gesamtanzahl der Mischkomponenten bzw. latenten Klassen entspricht dem Produkt der Kategorienanzahl der latenten

Klassenvariablen multipliziert mit der Anzahl der latenten Zustände hoch $t+1$ für $t=0,1,2,3,\dots,t$. Die Wahrscheinlichkeiten sind der gewichtete Durchschnitt der klassenspezifischen Wahrscheinlichkeitsdichte, wobei die vorangehenden Klassenzugehörigkeitswahrscheinlichkeiten oder Mischverteilungen als Gewichte dienen. In der *first-order* Markov-Annahme hängt eine Beobachtung zum Zeitpunkt t lediglich mit den Beobachtungen zu den Zeitpunkten $t+1$ und $t-1$ zusammen. Es resultieren vier unterschiedliche Wahrscheinlichkeiten, nämlich die Wahrscheinlichkeit einer latenten Klasse anzugehören und die Wahrscheinlichkeit eines latenten Zustandes zu Beginn gegeben die individuelle Klassenzugehörigkeit. Diese beiden Wahrscheinlichkeiten können von Kovariaten beeinflusst werden. Weiters gibt es die latente Veränderungswahrscheinlichkeit, welche die Wahrscheinlichkeit angibt, einen latenten Zustand zum Zeitpunkt t gegeben den latenten Zustand zum Zeitpunkt $t-1$ und Klassenzugehörigkeit, anzunehmen. Die Antwortwahrscheinlichkeit, ist die Wahrscheinlichkeit eine gewisse Antwort zum Zeitpunkt t gegeben den latenten Zustand zum Zeitpunkt t und die Klassenzugehörigkeit zu beobachten. Letztgenannte Wahrscheinlichkeiten können von *time-varying* Kovariaten abhängen. Das *mixture latent Markov model* beinhaltet somit die unbeobachtete Heterogenität (die interindividuelle Variabilität in den Veränderungspatterns) gemessen durch die zeitkonstante latente Variable, Autokorrelation (Tendenz, den gleichen latenten Zustand zwischen zwei benachbarten Beobachtungen beizubehalten) gemessen durch den Markov Veränderungsprozess erster Ordnung, und Messfehler oder Fehlklassifikation (fälschliche Veränderung aufgrund von Messfehlern) gemessen durch die ungenaue Beziehung zwischen zeitspezifischen latenten Zuständen und beobachteten Antworten. Tabelle 76 listet die Eigenheiten genannter Modelle auf.

4.3.7.6. Beispiele der *latent transition analysis*

Compton et al. (2008) nutzen eine LTA um die Stabilität latenter Klassen über die Zeit hinweg zu beschreiben. Der empirische Inhalt betrifft die Entwicklung der Fähigkeit lesen zu können. Es werden Beobachtungen über drei Zeitpunkte gemessen. In den vorgestellten Modellen sind jedoch lediglich zwei Zeitpunkte auf einmal enthalten. Um Aussagen zu Personeneigenschaften höherer Anfälligkeit zum Leseunvermögen machen zu können, werden Unterschiede verschiedener Indikatoren zum ersten Zeitpunkt zwischen den latenten Klassen gemessen, also nicht als Kovariate eingebracht, sondern im Anschluss daraus Mittelwerte berechnet. Weiters erfolgt eine verzerrende Kodierung, um das Auffinden unterschiedlicher Klassen künstlich zu erleichtern. So wird für die einzelnen Variablen lediglich ein Schnittpunkt definiert, welcher die beobachteten Variablen vorweg in die später aufgefundenen Gruppen

dichotomisiert. Die Autoren merken an, dass ein einzelner Schnittpunkt die Personen aufgrund von Messfehlern entlang dieses Schnittpunktes fluktuieren lässt und eine Instabilität der wiederholten Klassifikation verursacht. Mehrere wiederholte Messungen um die latenten Klassen zu formen, sollten die Stabilität jedoch verbessern. Zwei bzw. drei Indikatoren, welche dem theoretischen Konstrukt des Lesevermögens zugeteilt sind, werden zusammengefasst um die latenten Klassen zu formen. Das Modell geht von Messinvarianz der latenten Klassenindikatoren aus. Weiters werden jeweils zwei kategoriale latente Variablen, eine zu jedem Zeitpunkt, über multinomial logistische Regression miteinander verbunden, welche die Personen letztendlich in ‚reading disability (RD)‘ und ‚typical development (TD)‘ unterteilen. Für jedes Modell existieren über die Zeit hinweg betrachtet vier latente Klassen, zwei stabile und zwei instabile, nämlich RD->RD, TD->TD, RD->TD und TD->RD. Das Ergebnis liefert den Anteil an Individuen in jeder latenten Klasse zu Beginn und die latente Klassenwahrscheinlichkeit, welche die Wahrscheinlichkeit einer korrekten Zuteilung angibt, je näher bei eins desto besser. Weiters wird die Wechselwahrscheinlichkeit der Individuen zwischen den latenten Klassen in der Übergangswahrscheinlichkeitsmatrix ausgegeben. Diese enthält die geschätzten Wahrscheinlichkeiten der latenten Klassenzugehörigkeit zum nächsten Zeitpunkt, bedingt durch die latente Klassenwahrscheinlichkeit zum vorangegangenen Zeitpunkt.

Model name	Transition structure	Unobserved heterogeneity	Measurement error
I. Mixture latent Markov	yes	yes	yes
II. Mixture Markov	yes	yes	no
III. Latent Markov	yes	no	yes
IV. Standard Markov *	yes	no	no
V. Mixture latent growth	no	yes	yes
VI. Mixture growth	no	yes	no
VII. Standard latent class	no	no	yes
VIII. Independence*	no	no	no

*: This model is not a latent class model

Tabelle 76: Klassifikation latenter Klassenmodelle für longitudinale Analysen (Vermunt et al., 2008)

Kaplan (2008) stellt empirisch die unterschiedlichen Formen von Markov Ketten vor, um die Veränderung über die Zeit in Bezug auf den Entwicklungsstatus zu messen. Ausgehend vom *manifest* Markov *chain*-Modell, welches Veränderungen über die Zeit in kategorialen Variablen misst, werden Erweiterungen

zeigt welche Messfehler in den Antworten oder Abfolgen qualitativer Zustände erlauben. Das *mixture latent* Markov-Modell ist das allgemeinste, welches Messfehler in den kategorialen Antworten, als auch unbeobachtete Heterogenität in den Übergangswahrscheinlichkeiten erlaubt. Diese Annahme geht davon aus, dass unterschiedliche Subpopulationen mit eigenen zeitlichen Veränderungen existieren. Zwei Fälle können aus dem *mixture latent* Markov-Modell abgeleitet werden. In der LTA und den latenten Markov-Modellen, wird davon ausgegangen, dass es eine einzige Population mit einer Übergangswahrscheinlichkeit gibt und Messfehler in den kategorialen Antworten erlaubt sind. Daraus gehen zwei weitere Modelle hervor. LC-Modelle erlauben Messfehler in den kategorialen Antworten, die Daten werden aber nicht longitudinal verwendet. Werden die Daten longitudinal verwendet und perfekte Reliabilität zwischen den kategorialen Messungen angenommen, wird dies zum *manifest* Markov-Modell. Kaplan (2008) erläutert zuerst das *manifest* Markov-Modell. Es folgt das LC-Modell, welches Messfehler erlaubt und den Grundstein für *latent* Markov-Modelle bildet. Dann folgt das *latent* Markov-Modell und schließlich die LTA. Letztere erweitert das *latent* Markov-Modell, um multiple Indikatoren eines zugrundeliegenden latenten Zustandes verarbeiten zu können. Dieses Modell eignet sich besser für *stage-sequential development*, also stufenweise Entwicklungen. Dann wird die LTA um eine *mixture*-Komponente erweitert, was die Spezifikation eines *mover-stayer*-Modells erlaubt.

Das *manifest* Markov-Modell geht von kategorial beobachteten Daten aus. Es existiert nur eine Kette, welche zur Vorhersage des derzeitigen Individuenzustandes nur die vorangehende Beobachtung benötigt. Der beobachtete Anteil der Personen zum ersten Zeitpunkt beschreibt die anfängliche marginale Verteilung. Folgend werden die Übergangswahrscheinlichkeiten, gegeben die auf dem Modell basierten erwarteten Anteile an Personen zum vorhergehenden Zeitpunkt, errechnet. Diese stellen Wechselwahrscheinlichkeiten dar und bilden alle möglichen Übergänge zwischen zwei Zeitpunkten ab. Wird kein freies Wechseln zwischen Klassen zu jedem Zeitpunkt erlaubt, sondern nur in eine Richtung, jedoch nicht in die Rückrichtung, wird den Daten ein striktes Guttman-Antwortpatternverhalten unterstellt. Werden die Übergangswahrscheinlichkeiten konstant über die Zeit angenommen, spricht man vom *stationary* Markov *chain*, können sich diese über die Zeit unterscheiden, spricht man vom *nonstationary* Markov *chain*. Das *manifest* Markov-Modell geht von Antworten ohne Messfehler aus. Dies kann jedoch nicht immer angenommen werden. Deshalb sollten Messfehler des Markov-Modells explizit behandelt werden.

4.3.7.7. Die Messmodellebene

Das LC-Modell kann als zugrundeliegendes Messmodell für latente Markov-Modelle dienen. In der traditionellen LCA wird die Zugehörigkeit zu lediglich einer Klasse angenommen. Die beobachteten Antworten sind unabhängig voneinander, auch als lokale Unabhängigkeit bezeichnet. Latente Klassen stellen kategoriale Faktoren dar, welche aus den Antwortpatterns bestimmt werden. Dabei wird die Wahrscheinlichkeit einer bestimmten Antwort eines beobachteten Indikators gegeben die latente Klassenzugehörigkeit geschätzt. Kategoriale latente Variablen splitten folglich die Individuen in gemeinsam unabhängige Gruppen auf, die latenten Klassen. Der Anteil der Individuen in der Gruppe und die Antwortwahrscheinlichkeiten für die Items, bedingt durch die Zugehörigkeit zu der latenten Klasse, werden geschätzt. Dieses Modell ist für die latenten Markov-Modelle brauchbar, da eine konsistente latente Klassenstruktur über die Zeit spezifiziert werden kann.

Das *manifest* Markov-Modell geht davon aus, dass die beobachteten kategorialen Antworten perfekte reliable Messungen des wahren latenten Zustandes sind. Die beobachteten Antworten sind jedoch eher fehlerbehaftete Messungen eines nicht beobachtbaren latenten Zustandes. Tatsächlich sind aber die Übergänge zwischen den wahren latenten Zuständen von Interesse. Die LCA löst das Problem der Messfehler zu jedem Zeitpunkt durch latente Klassen. Zusammen mit *manifest* Markov-Modellen wird die Veränderung des latenten Levels über die Zeit gemessen, genannt *latent* Markov-Modell. Es werden Übergangswahrscheinlichkeiten auf latentem Level berechnet und es wurde von einer latenten Verteilung mit k latenten Zuständen ausgegangen. Die Verbindung der latenten Zustände zu den beobachteten Antworten läuft über Antwortwahrscheinlichkeiten. Es geht also um die Antwortwahrscheinlichkeit für eine bestimmte Kategorie gegeben die Zugehörigkeit zu einem latenten Status. Ziel ist die Messung der Veränderung auf dem latenten Level über die Zeit hinweg. Sind die Antwortwahrscheinlichkeiten alle eins, was eine perfekte Messung der latenten Variable bedeutet, reduziert sich dieses Modell zum *manifest* Markov-Modell.

Die LTA behandelt mehrere beobachtete Indikatoren. Multiple latente Variablen werden über multiple Indikatoren gemessen und gewinnen somit an Reliabilität und Validität. Deshalb werden multiple beobachtete kategoriale Indikatoren für die kategoriale latente Variable mit Markovketten-Modellen kombiniert. Die Verbindung von *multiple indicator categorical latent variable*-Modellen und Markovketten-Modellen bildet die Grundlage für die Analyse latenter Übergänge

stufenweiser dynamischer latenter Variablen. Multiple Indikatoren werden für die latente kategoriale Variable verwendet. Folglich ist das Messmodell für die kategoriale latente Variable das LC-Modell. Zu jedem Zeitpunkt definiert die latente Klassenzugehörigkeit den latenten Zustand. Modelliert wird ein spezieller zeitlicher Veränderungstyp des latenten Zustandes, genannt *stage-sequential development*. Der Prozess wird durch eine *stage-sequential dynamic latent-variable* repräsentiert. Verglichen mit latenten Markov-Ketten ist die LTA besser zur Messung konzeptioneller Zustandsentwicklungen geeignet.

4.3.7.8. Populationsunterschiede

Im *mixture latent* Markov-Modell wird nicht mehr von einer einzelnen Population, also einer einzigen Markovkette, latent oder auch nicht, und nur einem Set von Parametern ausgegangen, sondern von unterschiedlichen manifestierten Gruppen. Die Population besteht folglich aus einer finiten nicht beobachteten Mischung von Subpopulationen, charakterisiert durch qualitativ unterschiedliche Markovketten. So werden Modelle, welche auf Markovketten basieren, unter der Annahme einer *mixture*-Verteilung verbunden. Ein Spezialfall, das *mover-stayer*-Modell, erlaubt einer Gruppe zwischen Zuständen über die Zeit zu wandeln, *mover*, und der anderen Gruppe keine Übergänge zwischen den Zuständen, *stayer*. Die Anteile der Beobachtungen in den Markovketten werden errechnet und die Parameter sind unter der Bedingung der Zugehörigkeit zu einer speziellen Kette zu interpretieren. Gibt es nur eine Kette, wird dieses Modell zum *latent* Markov-Modell. Gibt es nur eine Kette und keine Übergangswahrscheinlichkeiten, wird es zum LC-Modell. Auf die zuletzt erörterte Modellgruppe wird später noch detailliert eingegangen.

Kaplan (2008) erläutert die Modellidentifizierung von manifesten Markovketten. Alle Parameter können direkt aus den beobachteten kategorialen Variablen ermittelt werden. In latenten Markovketten-Modellen mit nur einem Indikator müssen aus Identifizierungszwecken die Antwortwahrscheinlichkeiten als invariant über die Zeit angenommen werden. Dies bedeutet gleichbleibende Messfehler über die Zeit. Für vier oder mehr Zeitpunkte genügt die Invariantsetzung des ersten und des letzten Sets an Antworthäufigkeiten. Im *mixture*-Fall hängt die Identifizierung von der Anzahl an Ketten ab. Das *mover-stayer*-Modell ist mit einer einfachen Struktur und wenigen Parametern relativ gut schätzbar. Geschätzt wird üblicherweise mit ML über EM. Zur Ermittlung des Modellfits werden die beobachteten mit den aus dem Modell vorhergesagten Antwortanteilen verglichen. Dafür gibt es die Pearson- χ^2 -*statistic*, welche die beobachteten mit den erwarteten Zellenhäufigkeiten vergleicht, und die LR-*statistic*, welche asymptotisch χ^2 -verteilt ist. In Fällen wo größere Unstimmig-

keiten zwischen beiden Tests herrschen, kann von spärlich besetzten Zellen ausgegangen werden. Auch AIC und BIC werden zur Modellauswahl herangezogen, welche eine Bestrafungsfunktion für die Spezifizierung und Komplexitätstestung wenig sparsamer Modelle besitzen. BIC bestraft zusätzliche freie Parameter stärker, verglichen mit AIC. Das Modell mit dem niedrigsten AIC oder BIC wird bevorzugt.

4.3.7.9. *Nonstationary manifest Markov model - Kinodatensatz*

Kaplan (2008) zeigt eine empirische Abhandlung zuvor genannter Modelle, welche für den Kinopaneldatensatz adaptiert wird. Das *nonstationary manifest* Markov-Modell mit der Gesamtzufriedenheit liefert aufgrund der *most likely class membership* die Klassengrößen wie in Tabelle 77 aufgeführt. Klasse zwei ist deutlich größer als eins.

Zeitpunkt	Klasse	Absolute Anzahl	Anteil %	Zeitpunkt	Klasse	Absolute Anzahl	Anteil %
1	1	199	11,46	7	1	144	8,29
	2	1.538	88,54		2	1.593	91,71
2	1	166	9,56	8	1	178	10,25
	2	1.571	90,55		2	1.559	89,75
3	1	181	10,42	9	1	128	7,37
	2	1.556	89,58		2	1.609	92,63
4	1	160	9,21	10	1	157	9,04
	2	1.577	90,79		2	1.580	90,96
5	1	209	12,03	11	1	144	8,29
	2	1.528	87,97		2	1.593	91,71
6	1	160	9,21	12	1	119	6,85
	2	1.577	90,79		2	1.618	93,15

Tabelle 77: Klassengrößen - nonstationary manifest Markov Modell

Tabelle 78 zeigt die Übergangswahrscheinlichkeiten zu den jeweiligen Zeitpunkten. Es ist erkennbar, dass Klasse zwei relativ stabil über die Zeit ist, nämlich mit einer durchschnittlichen Wahrscheinlichkeit von über 90% in der Klasse zu verbleiben, 88,7% bis 93,8%. Personen in Klasse eins zeigen eine hohe Tendenz in Klasse zwei zu wandern, 62,3% bis 76,7%. Die Wahrscheinlichkeit in Klasse eins zu verbleiben liegt zwischen 23,3% und 37,7%. Die Wahrscheinlichkeit von Klasse zwei in Klasse eins zu wandern liegt zwischen 11,3% und 6,2%. Es kann allgemein die Aussage gemacht werden, dass die Übergänge zwischen den Klassen über die Zeit hinweg relativ stabil sind.

		Zeitpunkt 2	
		Klasse 1	Klasse 2
Zeitpunkt 1	Klasse 1	0,233	0,767
	Klasse 2	0,099	0,901

		Zeitpunkt 3	
		Klasse 1	Klasse 2
Zeitpunkt 2	Klasse 1	0,262	0,738
	Klasse 2	0,098	0,902

		Zeitpunkt 4	
		Klasse 1	Klasse 2
Zeitpunkt 3	Klasse 1	0,291	0,709
	Klasse 2	0,078	0,922

		Zeitpunkt 5	
		Klasse 1	Klasse 2
Zeitpunkt 4	Klasse 1	0,377	0,623
	Klasse 2	0,113	0,887

		Zeitpunkt 6	
		Klasse 1	Klasse 2
Zeitpunkt 5	Klasse 1	0,325	0,675
	Klasse 2	0,080	0,920

		Zeitpunkt 7	
		Klasse 1	Klasse 2
Zeitpunkt 6	Klasse 1	0,304	0,696
	Klasse 2	0,076	0,924

		Zeitpunkt 8	
		Klasse 1	Klasse 2
Zeitpunkt 7	Klasse 1	0,369	0,631
	Klasse 2	0,097	0,903

		Zeitpunkt 9	
		Klasse 1	Klasse 2
Zeitpunkt 8	Klasse 1	0,310	0,690
	Klasse 2	0,062	0,938

		Zeitpunkt 10	
		Klasse 1	Klasse 2
Zeitpunkt 9	Klasse 1	0,373	0,627
	Klasse 2	0,077	0,923

		Zeitpunkt 11	
		Klasse 1	Klasse 2
Zeitpunkt 10	Klasse 1	0,357	0,643
	Klasse 2	0,074	0,926

		Zeitpunkt 12	
		Klasse 1	Klasse 2
Zeitpunkt 11	Klasse 1	0,300	0,700
	Klasse 2	0,071	0,929

Tabelle 78: Transition probabilities

4.3.7.10. *Stationary manifest vs. stationary latent Markov model - Kinodatensatz*

Im Folgenden werden das manifeste Markov-Modell und das latente Markov-Modell unter Annahme einer stationären Markov-Kette verglichen. Die Übergangswahrscheinlichkeiten beider Modelle werden in Tabelle 79 verglichen, um zu sehen, ob die Miteinbeziehung der Messfehler in den manifestierten kategorialen Antworten Unterschiede in den Ergebnissen zur Folge hat. Die Übergangswahrscheinlichkeiten in den stationären Modellen sind über die Zeit hinweg homogen. Die Antwortwahrscheinlichkeiten im stationären manifesten und stationären latenten Markov-Modell sind invariant. Im manifesten Modell schwankt die Größe der ersten Klasse zwischen 119 und 209, die der zweiten zwischen 1.618 und 1.528. Im latenten Modell schwankt die Größe der ersten Klasse zwischen 18 und 48, die der zweiten zwischen 1.719 und 1.689. Hier ist zu bemerken, dass sich im latenten Modell die Klassengrößen bei 48 und 1.689 in Welle neun einpendeln und nicht mehr verändern. Bezüglich der Unterschiede in den Übergangswahrscheinlichkeiten ist zu bemerken, dass das latente Modell wesentlich stabilere Klassen zeigt, verglichen mit dem manifesten.

	Manifest Markov Chain		Latent Markov chain	
	1	2	1	2
1	0,313	0,687	0,917	0,083
2	0,085	0,915	0,004	0,996

Tabelle 79: Stationary manifest vs. stationary latent Markov Modell

4.3.7.11. *Cross-sectional latent class models - Kinodatensatz*

Nun wird von der Gesamtzufriedenheit abgegangen und auf Basis mehrerer Zufriedenheitsitems gerechnet. Zuerst werden jedoch LC-Modelle separat für jeden Zeitpunkt geschätzt. Diese Voruntersuchungen mit separaten *cross-sectional* LC-Modellen schlagen durch den besten Gesamtfitt und Erklärung der beobachteten Antworthäufigkeiten eine Klassenanzahl vor. Tabelle 80 zeigt eine Drei-Klassenlösung welche sich hier am besten bewährt.

Es ist wiederum deutlich eine sehr große Gruppe mit einer sehr hohen Wahrscheinlichkeit, in allen Items eine zufriedene Bewertung abzugeben, erkennbar. Weiters erwecken die relativ flachen Verläufe der Itemwahrscheinlichkeiten den Eindruck, dass die weitere Aufspaltung nach dem Wahrscheinlichkeitsgrad einer positiven Antwort abgestuft wird, also eine mittel zufriedene und eine weniger zufriedene.

Zeitpunkt	Latente Klasse	Antwortwahrscheinlichkeiten einer positiven Zufriedenheitsbewertung					Most Likely class membership
		Standort/ Erreichbarkeit	Filmangebot	Informationsangebot	Reservierungsmöglichkeiten	Erscheinungsbild/Image	
1	1	0,967	0,993	0,984	0,968	0,973	93,93
	2	0,430	0,638	0,234	0,237	0,000	2,19
	3	0,944	0,601	0,686	0,834	0,876	3,88
2	1	0,974	0,990	0,987	0,961	0,976	92,14
	2	0,713	0,832	0,540	0,478	0,000	3,54
	3	0,865	0,491	0,697	0,782	1,000	4,32
3	1	1,000	0,973	0,998	1,000	1,000	84,23
	2	0,660	0,198	0,316	0,793	0,425	3,25
	3	0,872	1,000	0,951	0,819	0,832	12,53
4	1	0,958	0,974	1,000	0,970	0,968	92,57
	2	0,230	0,457	0,385	0,674	0,321	2,96
	3	1,000	0,768	0,497	0,691	0,826	4,47
5	1	0,951	0,965	0,987	0,947	0,956	94,32
	2	0,284	0,803	0,758	0,859	0,000	2,60
	3	0,679	0,482	0,000	0,669	0,815	3,09
6	1	0,979	0,979	1,000	0,964	0,972	93,16
	2	0,610	0,000	0,000	0,277	0,000	7,43
	3	0,875	0,762	0,621	0,751	0,782	6,10
7	1	0,970	0,983	0,994	0,970	0,960	93,64
	2	0,593	0,510	0,463	0,392	0,000	2,04
	3	0,951	0,585	0,618	0,781	1,000	4,32
8	1	0,992	0,969	0,994	0,956	0,985	90,35
	2	0,800	0,433	0,000	0,694	0,469	2,38
	3	0,703	0,888	0,973	0,775	0,698	7,27
9	1	0,971	1,000	1,000	0,950	0,967	91,68
	2	0,384	0,488	0,329	0,385	0,206	1,73
	3	1,000	0,486	0,705	0,822	0,910	6,59
10	1	0,950	1,000	1,000	0,957	0,966	90,68
	2	0,908	0,522	0,725	0,894	1,000	5,31
	3	0,555	0,583	0,469	0,597	0,000	4,01
11	1	0,992	1,000	0,997	0,984	1,000	82,51
	2	0,884	0,855	0,922	0,900	0,769	14,91
	3	0,804	0,356	0,000	0,341	0,628	2,57
12	1	0,975	0,981	0,997	0,964	0,963	94,82
	2	0,021	0,015	0,000	0,000	0,000	0,65
	3	0,860	0,650	0,569	0,770	0,606	4,53

Tabelle 80: Drei Klassen LCA

Diagramm 128 zeigt jedoch, dass die weiteren beiden Gruppen nicht erkennbar voneinander ausdifferenziert sind. Dargestellt werden hier die Itemwahrscheinlichkeiten der fünf Items für alle Zeitpunkte und jeweils drei Klassen (5 Itemwahrscheinlichkeiten x 12 Zeitpunkte x 3 Klassen). Theoretisch müssten die Verläufe am oberen Rand, in der Mitte und am unteren Rand jeweils eine Gruppe von zwölf Verläufen bilden, nämlich zu zwölf Zeitpunkten.

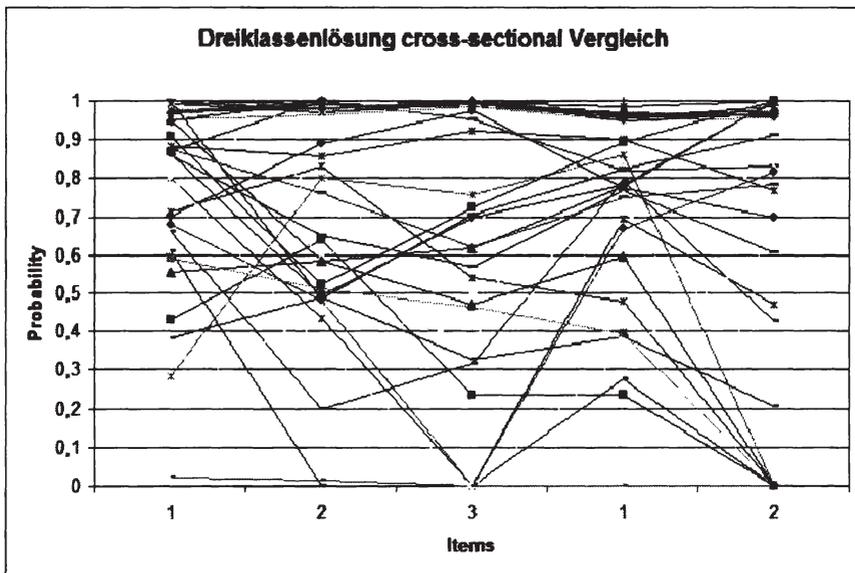


Diagramm 128: Drei-Klassenlösung cross-sectional Vergleich

4.3.7.12. Latent (hidden) Markov model - Kinodatenatz

Die LCA-Modelle erlauben aber nicht die Schätzung der Übergangswahrscheinlichkeiten, welche zwecks Analyse der Entwicklung über die Zeit fundamental sind. Im Folgenden wird deshalb schrittweise eine LTA aufgebaut. Zuerst wird ein *latent (hidden)* Markov-Modell mit nur einer Variablen, nämlich der Gesamtzufriedenheit, gerechnet, wie in Diagramm 129 dargestellt. In dieser Modellart wird nur ein binärer *outcome* verwendet. Obwohl jede kategoriale latente Variable nur einen latenten Klassenindikator besitzt, erlaubt das Modell die Schätzung eines Messfehlers der Gesamtzufriedenheit durch die Möglichkeit der Nicht-Übereinstimmung der latenten Klassenzugehörigkeit und beobachteten Antwort.

In folgendem Modell wird ein *first-order* Markov Prozess unterstellt, wobei die Übergangsmatrix über die Zeit gleich gehalten wird. So werden die *intercepts*

und die *first-order* Markov-Beziehungen aller latenten kategorialen Variablen über die Zeit gleich gehalten. Zusätzlich wird Messinvarianz angenommen. Tabelle 81 zeigt die konstante Veränderung zwischen den jeweiligen Zeitpunkten.

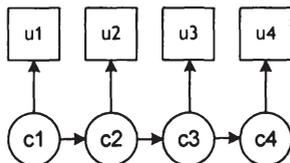


Diagramm 129: Hidden Markov Modell (Muthén und Muthén, 2007)

	1	2
1	0,993	0,007
2	0,102	0,898

Tabelle 81: Hidden Markov

Tabelle 82 gibt die Klassengrößen aus. Es zeigt sich, dass die Gruppe mit den positiven Zufriedenheitsantworten über die zwölf Monate hinweg größer wird. Sie steigt von 1.565 auf 1.604. Die Größe der kleineren unzufriedeneren Gruppe sinkt über die Zeit von 172 auf 133 Personen. Dies ist ein Indiz dafür, dass der stärkere Trend in Richtung positiver Zufriedenheitsaussagen geht.

Zeitpunkt	Klasse	Absolute Anzahl	Anteil %	Zeitpunkt	Klasse	Absolute Anzahl	Anteil %
1	1	1.565	90,10	7	1	1.596	91,88
	2	172	9,90		2	141	8,12
2	1	1.565	90,10	8	1	1.604	92,34
	2	172	9,90		2	133	7,66
3	1	1.575	90,67	9	1	1.606	92,46
	2	162	9,33		2	131	7,54
4	1	1.579	90,90	10	1	1.604	92,34
	2	158	9,10		2	133	7,66
5	1	1.581	91,02	11	1	1.604	92,34
	2	156	8,98		2	133	7,66
6	1	1.591	91,60	12	1	1.604	92,34
	2	146	8,41		2	133	7,66

Tabelle 82: Klassengrößen - Hidden Markov Modell

Die Übergangswahrscheinlichkeiten der Gesamtzufriedenheitsantworten bleiben über die Zeit hinweg gleich. Somit ergeben sich für den *item probability plot* in Diagramm 130, welcher die Antwortwahrscheinlichkeiten in den Gruppen abbildet, zwei horizontale Linien. Zwischen den Gruppen springen jedoch die Personen wie in vorheriger Tabelle aufgezeigt hin und her. Für die erste Gruppe ist es deutlich wahrscheinlicher eine positive Antwort auf die Gesamtzufriedenheitsfrage zu geben, verglichen mit der zweiten Gruppe. Betrachtet man die am häufigsten auftretenden Patterns aufgrund derer *most likely latent class membership*, nämlich 1 1 1 1 1 1 1 1 1 1 1 mit 1.554 Personen und das am zweihäufigsten auftretende Pattern, nämlich 2 2 2 2 2 2 2 2 2 2 2 mit 122 Personen, wird klar, dass die meisten Personen in Bezug auf die Gesamtzufriedenheit ein relativ stabiles Bild zeigen.

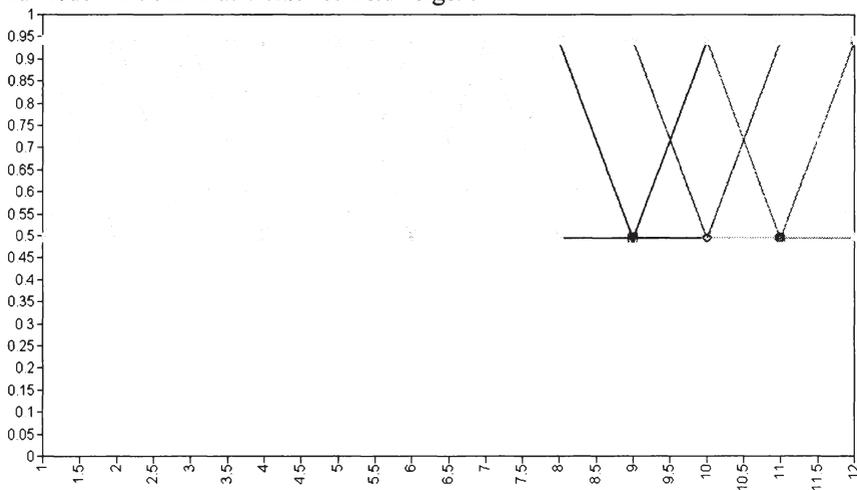


Diagramm 130: Item probability plot - hidden Markov-Modell

4.3.7.13. *Latent transition analysis* – Beispiele aus der Literatur

Collins et al. (2000) besprechen die LTA anhand eines empirischen Beispiels. Die meisten SEM gehen von kontinuierlichen latenten Variablen aus. LTA-Modelle verwenden diskrete latente Variable. Somit befinden sich die Personen nicht entlang eines latenten Kontinuums, sondern in latenten Kategorien. Auch die beobachteten Indikatoren werden als diskrete kategoriale Variable innerhalb der LC-Theorie betrachtet. Die Faktorenanalyse reproduziert die Korrelations- oder Kovarianzmatrix, die LCA reproduziert die Häufigkeiten der Antwortpatterns. Somit existieren Parameter, welche die Beziehung zwischen der latenten Variablen und den manifesten Variablen beschreiben. In der Faktorenanalyse repräsentieren diese die Faktorenladungen in Form von

standardisierten Regressionskoeffizienten. In LC-Modellen sind dies die Wahrscheinlichkeiten zwischen null und eins. Somit sind LC-Modelle Faktorenanalysen mit kategorialen Daten und kategorialen Faktoren. Geschätzt werden der Zustand und das Auftreten der Zustandsübergänge korrigiert um den Messfehler. LTA ist also eine Erweiterung der LC-Theorie für longitudinale Daten, wobei sowohl die Indikatoren, als auch die latente Variable kategorial sind. LTA erlaubt die Messung von Übergängen zwischen Zuständen. Verglichen mit dem LC-Modell sind die zusätzlichen Parameter die Verteilung der Zustandszugehörigkeit zu jedem Zeitpunkt und die Wahrscheinlichkeit in einen Zustand zu wandern, bedingt durch die vorangehende Zugehörigkeit zu einer bestimmten Klasse. Die Pfeile in Diagramm 131 stellen Übergänge von einem Zustand zu einem anderen dar, also eine Wechselwahrscheinlichkeit ungleich null, verglichen mit SEM, wo die Pfeile Pfadkoeffizienten ungleich null repräsentieren.

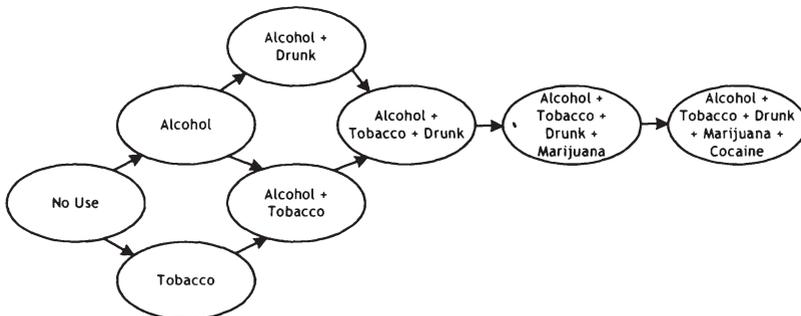


Diagramm 131: Stage-sequential developmental-Modell (Collins et al., 2000)

Zur Modellbewertung merken Collins et al. (2000) zur χ^2 -verteilten *goodness of fit*-Index Schätzung des EM Algorithmus an, dass bei vorliegender ausreichender Stichprobengröße die Kreuzvalidierung für die Modellselektion geeigneter ist, da die χ^2 -Verteilung nur schwer haltbar ist. Die Freiheitsgrade ergeben sich aus der Anzahl der möglichen Antwortpatterns, welche für gewöhnlich größer ist, verglichen mit der beobachteten minus der Anzahl an unabhängig geschätzten Parametern, minus eins. In ihrer Modellbewertung merken sie jedoch an, dass ein extrem kleinerer Wert des *goodness of fit*-Index verglichen mit der Anzahl der Freiheitsgrade, als Hinweis auf einen guten Fit dient.

In obigem Diagramm führen sie eine exogene statische latente Variable ein, nämlich das Geschlecht, welches die Population in latente Klassen teilt. Um die Modelle einfacher interpretierbar zu machen, werden beispielsweise die Parameter des dynamischen Teiles gleichgesetzt. So wird die Bedeutung der

latentem Zustände über die Zeit hinweg gleich gehalten. Auch Beschränkungen oder Gleichsetzungen über Gruppen hinweg können notwendig sein. Auch in der Übergangsmatrix können Restriktionen eingeführt werden, um die Richtung der Wechselwahrscheinlichkeit zu lenken. Wird eine Beschränkung eingeführt, welche einen Wechsel nur in eine Richtung, nämlich ‚vor‘, aber nicht ‚zurück‘ erlaubt, werden die unteren Dreiecksmatrixelemente der Übergangsmatrix auf null fixiert. Der statische Teil wird verwendet, um die latenten Zustände zu beschreiben. Je näher die Werte an den Grenzen null und eins liegen, desto stärker ist die Beziehung zwischen den latenten Zuständen und der manifesten Variablen. Zusätzlich zu den Zugehörigkeitswahrscheinlichkeiten zu einem latenten Zustand, bedingt durch Geschlecht zum ersten Zeitpunkt, werden die 95%-igen Konfidenzintervalle angeführt, was Aussagen zu den unterschiedlichen Ausgangszuständen zwischen den Geschlechtern zulässt. Die Wechselwahrscheinlichkeiten liefern zwischen den beiden Geschlechtern in sämtlichen Parametern Überschneidungen und erst die 90%-igen Konfidenzintervalle lassen für einige Parameter geschlechterspezifisch unterschiedliche Interpretationen zu.

Collins et al. (2000) gehen auch auf die Vorteile der LTA ein. Kreuztabellen über alle beobachteten Antwortpatterns führen zu riesigen nicht mehr interpretierbaren Kontingenztabellen. Zusätzlich misst die LTA eine Fehlerstruktur, wobei fragwürdige Antwortpatterns ihren passenden latenten Zuständen zugeordnet werden können. Die Veränderung wird in Form von Wechselwahrscheinlichkeiten ausgedrückt und stellt somit die Größe der Wechselwahrscheinlichkeit dar, als auch das finale Stadium. Weiters werden die Wechselwahrscheinlichkeiten bedingt durch die vorangehende Klassenzugehörigkeit geschätzt. Diese Miteinbeziehung ist für longitudinale Analysen sehr wichtig, da sie den Effekt, welchen der Startpunkt auf die Veränderung hat, miteinbezieht. So kann beobachtet werden, ob Personen in unterschiedlichen Zuständen zum ersten Zeitpunkt, sich schneller oder langsamer durch die Zustandssequenzen bewegen. Auch die Möglichkeit der Individuen unterschiedliche Wege zu nehmen ist vorteilhaft.

Nylund (2006, 2007) beschreibt die stufenweise Entwicklung simpler Modelle bis hin zu einer LTA erster Ordnung mit Kovariaten, einer *mover-stayer* latenten Variablen und einem *distal outcome*, um *stage-sequential* Wachstumshypothesen zu testen. Zuerst wird das Messmodell bestimmt und die Fit-Informationen verglichen. Um Hinweise auf Veränderungen aufzudenken, werden über die Zeit *cross-sectional*-Kreuztabellen der Klassenzugehörigkeiten erstellt. Dann wird das LTA-Modell erstellt und Wechselwahrscheinlichkeitsspezifikationen in Form von beispielsweise stationären Prozessen oder Effekten höherer Ordnung

definiert und um Kovariate erweitert. Später wird eine latente Kovariate in Form einer latenten Klassenvariable höherer Ordnung zur Beschreibung un-beobachteter heterogener Entwicklung eingeführt, nämlich eine *mover-stayer*-Variable, welche zusätzlich Stabilität der Klassenzugehörigkeit ermöglicht.

LTA baut auf zwei unterschiedlichen Analyseverfahren auf, einerseits LCA und andererseits autoregressive Modelle, im speziellen Markov-Modelle. Die LCA als *latent variable mixture model* wird in der LTA als Messmodell verwendet, um Gruppen für die einzelnen Zeitpunkte zu beschreiben. Die autoregressive Komponente beschreibt den Übergang zwischen den Klassen über die Zeit. Für das *mover-stayer*-Modell wird eine zusätzliche latente Kovariate aufgesetzt, die *second-order mover-stayer*-Variable. Diese beschreibt die heterogene Entwicklung. LTA fällt in die Gruppe der SEM und besteht aus einem Mess- und einem Strukturmodell. Zuerst werden deshalb LCA-Modelle geschätzt. Nachdem die Klassen identifiziert wurden, wird die LTA verwendet um Veränderungen bzw. Veränderungspatterns zu studieren. LTA verwendet einen *person-centered model-based*-Ansatz um Klassen zu identifizieren. Somit benötigt sie keine externen Kriterien oder *cut points* für die Klassen. Die LTA modelliert Entwicklungspatterns zur Klassifizierung über die Zeit. Sie ist ein autoregressives Modell, welches Veränderungen in kategorialen latenten Variablen misst. Somit beschreibt LTA Veränderungen von *outcomes*, wobei diese nicht direkt beobachtet werden. Die *outcomes* in der LTA sind latent und durch ein Set an beobachteten Variablen abgebildet. Der *outcome* der LTA ist wiederum eine kategoriale Variable.

Im Vergleich zu den Wachstumsmodellen werden in den autoregressiven Modellen die wiederholt gemessenen *outcomes* nicht über die Wachstumsfaktoren verbunden, sondern direkt. Deshalb gibt es keine Annahmen bezüglich der Entfernung zwischen zwei Zeitpunkten. Sie betrachten die benachbarten Beziehungen zwischen den *outcomes* der Individuen. Die Korrelationen der *outcomes* sinken für gewöhnlich, je weiter die Zeitpunkte auseinander liegen, also je weiter sie von der Diagonale der Korrelationsmatrix entfernt sind. Im autoregressiven Modell existiert zu jedem Zeitpunkt eine Regression auf eine oder mehr davor liegende Zeitpunkte. Werden die Koeffizienten über die Zeit hinweg gleich gehalten, beschreibt dies einen stationären Prozess. Autoregressive Modelle werden jedoch oft für diskontinuierliche Verläufe verwendet. In den Wachstumsmodellen wird für die binären oder geordnet polytomen kategorialen *outcomes* eine kontinuierliche zugrunde liegende Verteilung angenommen, wobei die Kategorien die Schnittpunkte dieser kontinuierlichen Variablen darstellen. In autoregressiven Modellen beschreiben bedingte

Wahrscheinlichkeiten die Veränderungen zwischen den Kategorien in Form einer multinomial logistischen Regression, wobei die Variable zum Zeitpunkt t auf die Variable zum Zeitpunkt $t-1$ regressiert wird.

4.3.7.14. Transition probability matrix - Spezifizierung

Nylund (2007) beschreibt die Berechnung der Übergangsmatrixfelder mit zwei kategorialen Variablen und jeweils drei Kategorien. Dies kommt der Veränderung von drei Klassen auf drei neue Klassen zu einem späteren Zeitpunkt gleich. Um die Wechselwahrscheinlichkeit in eine Kategorie zum Zeitpunkt t , gegeben die individuelle Klassenzugehörigkeit zum Zeitpunkt $t-1$, zu bestimmen, müssen die Werte aus Tabelle 83 berechnet werden.

		C_t		
		1	2	3
C_{t-1}	1	$P(C_t=1 C_{t-1}=1)$	$P(C_t=2 C_{t-1}=1)$	$P(C_t=3 C_{t-1}=1)$
	2	$P(C_t=1 C_{t-1}=2)$	$P(C_t=2 C_{t-1}=2)$	$P(C_t=3 C_{t-1}=2)$
	3	$P(C_t=1 C_{t-1}=3)$	$P(C_t=2 C_{t-1}=3)$	$P(C_t=3 C_{t-1}=3)$

Tabelle 83: Transition matrix (Nylund, 2007)

Die Notationen für die drei Klassen zu den jeweiligen Zeitpunkten werden in Formel 26 angeschrieben. β stellen die logistischen Regressionskoeffizienten dar. Sie repräsentieren die Veränderungen der *logits* in Bezug auf die Veränderung einer Einheit in der unabhängigen Variablen. Wird mit Dummy-Variablen gerechnet, beschreiben diese den Unterschied in den *log odds* zwischen den Individuen der Dummyvariablen-Kategorie und der Referenzkategorie. Dieser Wert ist der natürliche Logarithmus der *odds ratio*, der Dummy-Variablen Kategorie versus der Referenzkategorie anzugehören. Um den logistischen Regressionskoeffizienten interpretieren zu können, muss dieser in die *odds ratio* konvertiert werden. Dazu wird der Exponent des Koeffizienten berechnet, z.B. $e^{\beta_{12}}$ bzw. $\exp(\beta_{12})$. Dieser ist die *odds ratio* der Klasse zwei zum Zeitpunkt t versus Klasse drei (die Referenzklasse), für jene, welche in Klasse eins zum Zeitpunkt $t-1$, verglichen mit jenen, welche in Klasse drei (der Referenzklasse) waren.

Die posterioren Klassenwahrscheinlichkeiten, welche die Zugehörigkeitswahrscheinlichkeiten jedes Individuums zu jeder Klasse des gefitteten Modells beschreiben, werden über die beobachteten Antwortpatterns der Items berechnet. Sie sind eine Funktion der geschätzten Itemwahrscheinlichkeiten und der geschätzten Prävalenz jeder latenten Klasse. Jedes Individuum wird der Klasse

mit der höchsten posterioren Klassenzugehörigkeitswahrscheinlichkeit zugeordnet.

$$\begin{aligned}
 P(C_t = 1 | C_{t-1} = 1) &= \frac{\exp(\alpha_1 + \beta_{11})}{\exp(\alpha_1 + \beta_{11}) + \exp(\alpha_2 + \beta_{21}) + 1} \\
 P(C_t = 2 | C_{t-1} = 1) &= \frac{\exp(\alpha_2 + \beta_{12})}{\exp(\alpha_1 + \beta_{11}) + \exp(\alpha_2 + \beta_{21}) + 1} \\
 P(C_t = 3 | C_{t-1} = 1) &= \frac{\exp(\alpha_3 + \beta_{13})}{\exp(\alpha_1 + \beta_{11}) + \exp(\alpha_2 + \beta_{21}) + 1} \\
 P(C_t = 1 | C_{t-1} = 2) &= \frac{\exp(\alpha_1 + \beta_{21})}{\exp(\alpha_1 + \beta_{12}) + \exp(\alpha_2 + \beta_{22}) + 1} \\
 P(C_t = 2 | C_{t-1} = 2) &= \frac{\exp(\alpha_2 + \beta_{22})}{\exp(\alpha_1 + \beta_{12}) + \exp(\alpha_2 + \beta_{22}) + 1} \\
 P(C_t = 3 | C_{t-1} = 2) &= \frac{\exp(\alpha_3 + \beta_{23})}{\exp(\alpha_1 + \beta_{12}) + \exp(\alpha_2 + \beta_{22}) + 1} \\
 P(C_t = 1 | C_{t-1} = 3) &= \frac{\exp(\alpha_1)}{\exp(\alpha_1) + \exp(\alpha_2) + 1} \\
 P(C_t = 2 | C_{t-1} = 3) &= \frac{\exp(\alpha_2)}{\exp(\alpha_1) + \exp(\alpha_2) + 1} \\
 P(C_t = 3 | C_{t-1} = 3) &= \frac{\exp(\alpha_3)}{\exp(\alpha_1) + \exp(\alpha_2) + 1}
 \end{aligned}$$

Formel 26: Transition Matrix Felder

4.3.7.15. Latente Ebenen

Wachstumsmodelle können auch über ein Modell höherer Ordnung die Veränderung in einer latenten kategorialen Variablen messen, soweit diese geordnet ist. Für die Veränderungsmessung latenter Klassenvariablen ist dies jedoch nicht gebräuchlich, sondern es werden autoregressive Modelle verwendet. In LTA ist das Messmodell meist ein LCA-Modell. Markovketten werden verwendet, um den Wechsel in einen neuen Zustand zu beschreiben. Markov-Modelle können erweitert werden, z.B. hinsichtlich der nicht notwendigen perfekten Messung der *outcomes* oder durch Erlaubnis einer Population höherer Ordnung, wobei unterschiedliche Wechselwahrscheinlichkeiten möglich sind. Somit verbindet die LTA eine *cross-sectional*-Messung der kategorialen latenten Variablen und der longitudinalen Beschreibung der Veränderung in den Kategorien der latenten Variablen. In den LCA-Modellen ist diese zugrundeliegende latente Variable kategorial. Die beobachteten Indikatoren sind im vorliegenden Fall ebenfalls kategorial, nämlich binär recodiert auf eins, zwei und

drei als ‚zufrieden (1)‘ und vier, fünf und sechs als ‚nicht zufrieden (0)‘. Sind die manifesten Indikatoren kontinuierlich, wird die LCA auch *latent profile analysis* genannt. Es werden Itemwahrscheinlichkeitsparameter bedingt durch die Zugehörigkeit zu einer latenten Klasse und durch die Klassenwahrscheinlichkeitsparameter, welche die relative Häufigkeit der Klassenzugehörigkeit für die Population angeben, geschätzt. Nylund (2007) zeigt die unterschiedliche Spezifizierung einer geordneten versus nicht geordneten Klassenlösung in Diagramm 132. Bei letzterer sind Überschneidungen möglich.

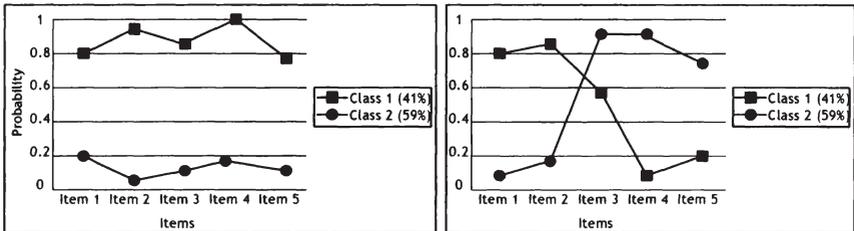


Diagramm 132: LCA item probability plots geordneter (links) vs. nicht-geordneter (rechts) binärer Daten (Nylund, 2007)

Nylund (2007) beginnt den Aufbau einer LTA über die LCA, um eine Entscheidung über die Anzahl der Klassen treffen zu können. Die *loglikelihood*-Werte dürfen nicht für den LRT zwischen den genesteten Modellen k versus $k-1$ benutzt werden. Die $k-1$ Lösung ist nämlich ein Spezialfall der k Lösung, in welcher die Wahrscheinlichkeit für eine der Klassen auf null gesetzt wird, was bedeutet, dass diese auf die Grenze des Parameterraumes fixiert wird und somit der Differenzwert zwischen den *loglikelihoods* nicht mehr χ^2 -verteilt ist. Herangezogen werden BLRT, welcher bei Nicht-Signifikanz das schmalere Modell mit einer Klasse weniger bevorzugt und bei Signifikanz das komplexere, BIC, welcher die Anzahl der Parameter und die Stichprobengröße bestraft, und ABIC, welcher Modelle mit einer größeren Stichprobengröße weniger bestraft. Nylund (2007) verwendet zur Klassenanzahlbestimmung zusätzlich *latent class factor analysis* (LCFA) und *factor mixture analysis* (FMA), beide hybride Modelle. LCFA wird für nicht normale Faktorverteilungen herangezogen, in welchen die latenten Klassen als Unterstützungspunkte für die Verteilung dienen und es keine *within-class*-Variation des Faktors gibt. FMA erlaubt *within-class*-Variation der Faktoren, wobei Invarianz der Messparameter für die Faktoren über die Klassen spezifiziert werden kann oder auch die Anzahl der Faktoren innerhalb der latenten Klassen unterschiedlich sein kann. Auch Faktorenanalysen werden durchgeführt. Tabelle 84 gibt einen Modellüberblick.

	Continuous Latent Variables	Categorical Latent Variables	Hybrids
Cross-Sectional Models	Factor analysis, SEM	Regression mixture analysis, Latent class analysis	Factor mixture analysis
Longitudinal Models	Growth analysis (random effects)	Latent transition analysis, Latent class growth analysis	Growth mixture analysis

Tabelle 84: Modellüberblick (Muthén, 2006)

4.3.7.16. Modellfitvergleiche der *latent class analysis* - Kinodatensatz

Um den Fit der Messmodelle zu vergleichen, werden Plots der Eigenwerte zur Bestimmung der Faktorenanzahl herangezogen. Für die LCA-Vergleiche werden BIC und BLRT und für die hybriden Messmodelle BIC, *log likelihood* und die Anzahl der Parameter zum Fitvergleich der genesteten Modelle herangezogen. Um die Gefahr, dass die Lösung in ein lokales Maximum konvergiert, zu senken und die Replikation gleicher *likelihood*-Werte zu erhöhen, wird eine hohe Anzahl an Zufallsstartwerten verwendet. Das am besten passende Messmodell wird dann für die longitudinalen Lösungen verwendet, um dessen Stabilität über die Zeit hinweg zu prüfen. LCFA mit niedrigerem BIC werden bevorzugt. Ist das Sinken des *loglikelihood*-Wertes bei Steigung der Anzahl der Parameter gering, deutet dies auf das schmalere Modell hin. Eine zu geringe Verbesserung des *loglikelihood*-Wertes der FMA, verglichen mit der LCA, deutet auf die Nicht-Notwendigkeit der *within-class*-Variation der FMA hin. Eine geringe Fitverbesserung zwischen LCA und LCFA deutet auf eine Nicht-Notwendigkeit der *within-class*-Variation der LCFA hin. FA ermöglicht ein Dimensionskontinuum, LCA trifft eine Klassifizierung in unterschiedliche Klassen. Im Kinobeispiel werden Vergleiche zwischen LCA unterschiedlicher Anzahl an Klassen zu den einzelnen Zeitpunkten aus Tabelle 85 herangezogen.

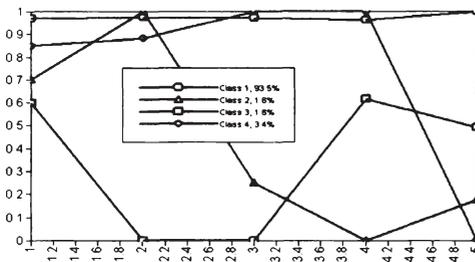
Achtmal wird das Drei-Klassenmodell bevorzugt, zweimal das Zwei-Klassenmodell und zweimal das Vier-Klassenmodell. Die fett markierten Lösungen sind die vom Modellfit her passendsten für die unterschiedliche Anzahl an Klassen.

Time point	Model	LogL	BIC	Free Estimated parameters	BLRT	Optimal # of classes		
1	LCA4c	-1.194.798	2.556.491	23	0.0400			4
1	LCA5c	-1.189.444	2.589.320	29	0.0938			
2	LCA3c	-1.260.070	2.643.449	17	0.0000		3	
2	LCA4c	-1.254.223	2.675.276	23	0.1176			
3	LCA4c	-1.316.402	2.800.652	23	0.0000			
3	LCA5c	-1.312.501	2.836.637	29	0.1875			
4	LCA3c	-1.409.202	2.942.187	17	0.0400		3	
4	LCA4c	-1.403.386	2.974.243	23	0.0732			
5	LCA3c	-1.500.752	3.124.968	17	0.0000		3	
5	LCA4c	-1.496.891	3.160.823	23	0.3750			
6	LCA2c	-1.190.445	2.460.135	11	0.0000	2		
6	LCA3c	-1.185.863	2.494.197	17	0.1765			
7	LCA3c	-1.105.527	2.333.219	17	0.0400		3	
7	LCA4c	-1.101.786	2.368.854	23	0.3750			
8	LCA3c	-1.191.678	2.506.373	17	0.0128		3	
8	LCA4c	-1.187.582	2.541.599	23	0.3750			
9	LCA3c	-1.115.144	2.351.838	17	0.0000		3	
9	LCA4c	-1.110.521	2.385.491	23	0.2000			
10	LCA3c	-1.416.793	2.957.568	17	0.0000		3	
10	LCA4c	-1.411.032	2.989.802	23	0.0896			
11	LCA3c	-1.169.149	2.460.463	17	0.0000		3	
11	LCA4c	-1.164.471	2.494.224	23	0.3333			
12	LCA2c	-1.006.168	2.090.653	11	0.0000	2		
12	LCA3c	-1.001.383	2.123.800	17	0.6667			

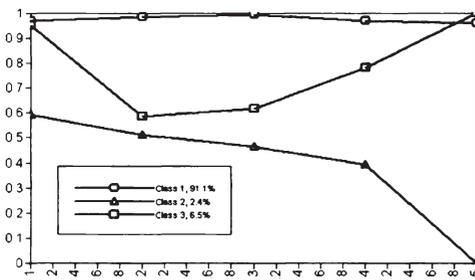
Tabelle 85: LCA Modellvergleich

Zur Interpretation der latenten Klassen werden Plots der bedingten Itemwahrscheinlichkeiten, *item probability plots*, herangezogen. Diese stellen Werte dafür dar, ob eine Person ein Item in einer Klasse bestätigt, also als zufriedenstellend beurteilt. Sie werden zur Beschreibung der Klassen herangezogen werden. Diagramm 133 vergleicht die fett markierten Lösungen unterschiedlicher Klassenanzahl. Rechts befinden sich jeweils die Zugehörigkeitswahrscheinlichkeiten der einzelnen Klassenlösungen. Die Linien geben die *profiles* an. Oberste Graphik zeigt die beste Vier-, mittlere die beste Drei- und letzte die beste Zwei-Klassenlösung. Klasse eins hat in allen drei unterschiedlichen Lösungen in allen fünf Items die höchste Wahrscheinlichkeit einer

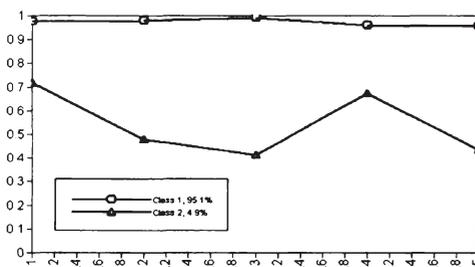
positiven Zufriedenheit und umfasst in allen drei Lösungen über 90% der Probanden, in der Zwei-Klassenlösung 95,1%, in der Drei-Klassenlösung Lösung 91,1% und in der Vier-Klassenlösung 93,5%. Die kleineren Klassen zeigen unterschiedliche Itemantworten. Allgemein ist die Zuordnung zu den latenten Klassen als zufriedenstellend zu bewerten, da abgesehen von zwei Zuordnungswahrscheinlichkeiten alle über 90% liegen, was auf eine ziemlich trennscharfe Zuordnung hindeutet. Ausreißer sind drei der Drei-Klassenlösung mit 74,1% und Cluster zwei der Zwei-Klassenlösung mit 86,3%. Das häufigste Antwortpattern ist 1 1 1 1 1, 945 beobachtete und 920,56 geschätzte in der Zwei-Klassenlösung mit 1.417 Probanden, 801 bzw. 787,47 in der Drei-Klassenlösung mit 1.321 und 745 bzw. 726,05 in der Vier-Klassenlösung mit 1.236.



	1	2	3	4
1	0.997	0.001	0.001	0.001
2	0.076	0.924	0.000	0.000
3	0.069	0.000	0.917	0.014
4	0.000	0.057	0.000	0.943



	1	2	3
1	0.960	0.004	0.035
2	0.062	0.935	0.002
3	0.248	0.011	0.741



	1	2
1	0.983	0.017
2	0.137	0.863

Diagramm 133: Item probability plots

Eins auf der x-Achse steht für die Bewertung Standort/Erreichbarkeit, zwei für Filmangebot, drei für Informationsangebot, vier für Reservierungsmöglichkeiten und fünf für Erscheinungsbild/Image. In der Zwei-Klassenlösung tendiert die kleinere Gruppe mit 4,9% der Probanden zu zufriedeneren Antworten in den Items Standort/Erreichbarkeit und Reservierungsmöglichkeiten und zu leicht schlechteren Zufriedenheitsantworten in den restlichen drei Items. In der Drei-Klassenlösung tendiert die Gruppe mit 2,4% zu schlechteren Werten, vor allem was das Erscheinungsbild/Image betrifft. Die etwas größere Gruppe mit 6,5% zeigt zufriedene Werte für Standort/Erreichbarkeit und Erscheinungsbild/Image. In der Vier-Klassenlösung sind die Zufriedenheitswerte der Gruppe mit 3,4% sehr hoch, hingegen der Wert für Erscheinungsbild/Image sehr niedrig. In der Gruppe mit 1,6% liegt die Wahrscheinlichkeit einer positiven Antwort auf Film- und Informationsangebot bei null Prozent, die Wahrscheinlichkeiten einer positiven Beantwortung der restlichen Items liegt im Mittelfeld. Die Gruppe mit 1,6% tendiert zu positiven Antworten für Standort/Erreichbarkeit und Filmangebot, nicht jedoch für die restlichen drei Items.

Tabelle 86 liefert eine Kurzbeschreibung der Gruppen.

Lösung	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Zwei Cluster	Zufrieden	Mittelmäßig zufrieden		
Drei Cluster	Zufrieden	Mittelmäßig zufrieden, schlechtes Image (Gefühlsebene)	Mittelmäßig zufrieden, standortbezogen sehr zufrieden	
Vier Cluster	Zufrieden	Zufrieden, nicht zufrieden vor Ort	Mittelmäßig zufrieden, nicht mit Angebot zufrieden	Zufrieden, mit Image nicht zufrieden

Tabelle 86: Clusterprofiles

4.3.7.17. *Partial measurement invariance vs. full measurement invariance vs. full measurement noninvariance*

Auch bezüglich der Invarianz des Messmodells können Spezifikationen getroffen werden. Eine gleichbleibende Anzahl an Klassen zu jedem Zeitpunkt legt Invarianz nahe, welche Gleichheit der Parameter des Messmodells voraussetzt, im speziellen die bedingten Itemwahrscheinlichkeiten der LCA-Variablen bzw. der Faktorladungen der Faktoren. In der LTA mit LCA sind die Messmodellparameter die bedingten Itemwahrscheinlichkeiten, welche für jede Klasse zu den unterschiedlichen Zeitpunkten geschätzt werden. Dafür stehen drei Levels an Invarianzmöglichkeiten zur Verfügung. *Full measurement invariance*

geht von invarianten bedingten Itemwahrscheinlichkeiten über alle Zeitpunkte aus. So wird die Bedeutung der Klassen durch gleichbleibende Klassenanzahl und –profile zu allen Zeitpunkten beibehalten. *Partial measurement invariance* geht von Gleichheit einiger Messparameter über die Zeit aus, wobei andere nicht beschränkt werden. Gleichheitsbeschränkungen der Itemwahrscheinlichkeiten für ein einzelnes Item über die Zeit innerhalb einer speziellen Klasse, bis hin zu Invarianz aller Items innerhalb einer Klasse über die Zeit, etc. sind möglich. Für drei Zeitpunkte, vier beobachtete Items und drei Klassen gibt es 36 Messparameter (4 Items \times 3 Klassen \times 3 Zeitpunkte). *Full measurement noninvariance* trifft keine Annahmen hinsichtlich der Gleichheit der Messparameter, also freie Schätzung über Klassen und Zeit. Die Klassen können sich über die Zeit hinweg wandeln und eine variierende Anzahl an Klassen über die Zeit ist erlaubt. Für den Vergleich zwischen den Invarianzmöglichkeiten bietet sich LRT an. Weniger restringente Lösungen werden mit jenem Modell, welches von gleicher Klassenanzahl und gleichen Klassentypen ausgeht, jedoch keine Restriktionen hinsichtlich der Klassengrößen hat, verglichen. Erst dann werden longitudinale, z.B. autoregressive Beziehungen, eingeführt. Beim Vergleich zwischen *complete measurement invariance* und *complete measurement noninvariance* bevorzugt ein signifikanter Wert die *complete measurement noninvariance*-Spezifikation.

Im Kinobeispiel ist kein Modell unter der *full measurement noninvariance*-Annahme schätzbar. Weiters wird aus Gründen der einfacheren und sinnvolleren Interpretation und im Hinblick auf die Vorbereitung des *mover-stayer*-Modells, *full measurement invariance* für sämtliche folgenden Modelle angenommen. Schon bei einem Modell über nur sechs Zeitpunkte, der beschränkten Annahme stationärer Übergangswahrscheinlichkeiten und nur zwei Klassen, ist die Anzahl an freien Parametern mit 1.922 höher als die Anzahl an Probanden mit 1.737. Selbiges Modell bei Annahme nicht-stationärer Wechselwahrscheinlichkeiten enthält 1.931 freie Parameter.

4.3.7.18. *First-order* vs. Autokorrelationen höherer Ordnung (*second-order*, *third-order*, ...)

Die Zusammenhänge zwischen den Zeitpunkten werden typischerweise *first-order* spezifiziert, also als autoregressive Struktur erster Ordnung. Die Markov-Struktur geht dabei von einer Abhängigkeit zwischen zwei benachbarten Zeitpunkten aus. Jede Variable X ist bedingt unabhängig von deren Nicht-Nachbarn gegeben die Nachbarn. Die generelle Notation kann wie in Formel 27 angeschrieben werden (Edwards, 2000) und es gilt für alle $X \in V$,

$$X \perp V \setminus \{X \cup bd(X)\} | bd(X)$$

Formel 27: Markov-Struktur

Anstatt der Variablen werden hier Zeitpunkte verwendet. Effekte höherer Ordnung anstatt der autoregressiven Spezifizierung erster Ordnung zwischen den latenten Variablen sind ebenfalls denkbar. C_1 , C_2 , und C_3 sind die latenten Variablen zu den drei Zeitpunkten und die u 's die manifesten beobachteten Items. Im Vergleich zu Diagramm 134, wird in Diagramm 135 zusätzlich zum direkten Einfluss von C_1 auf C_3 der indirekte Einfluss über C_2 miteinbezogen. Dies führt zu autoregressiven Strukturen höherer Ordnung. Zum Vergleich zwischen *first-* und *second-, third-* oder *higher-order* Modellen kann LRT verwendet werden. Da eine Verkettung höherer Ordnung als *first-order* Markov-Modell zu extrem umfangreichen Schätzungen führt und weiters eine enorme Anzahl an weiteren möglichen Verkettungen theoretisch denkbar ist, werden in sämtlichen folgenden Modellen die Messmodelle zu den einzelnen Zeitpunkten lediglich mit den benachbarten Messmodellen verkettet, was zu einer Markovkette erster Ordnung führt.

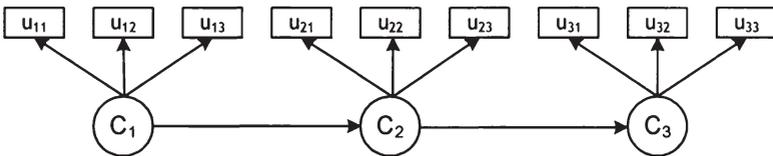


Diagramm 134: LTA – first-order Effekt (Nylund, 2007)

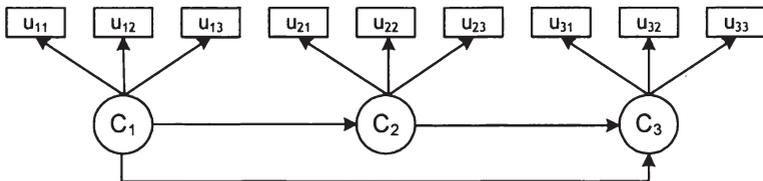


Diagramm 135: LTA - first- und second-order Effekt (Nylund, 2007)

4.3.7.19. Modellvergleichsmöglichkeiten und Einschränkungen der transition matrix

Auch Beschränkungen hinsichtlich der Übergangswahrscheinlichkeitsmatrix können spezifiziert werden. Stationäre Übergänge führen zu gleichbleibenden Übergangswahrscheinlichkeiten zwischen den Zeitpunkten. Um den Fit der Modelle zu vergleichen ist die Häufigkeitstabellen- χ^2 -statistic nach Pearson oder LR-basiert für LTA Modelle nicht geeignet, da die χ^2 -Verteilung bei einer

großen Anzahl an leeren Zellen nicht gut approximiert werden kann. Um genestete LTAs zu vergleichen wird LRT verwendet. Genestet sind Modelle unterschiedlicher Wechselwahrscheinlichkeiten oder Messparameter. LRT wird hier zum Vergleich zwischen stationär und nicht stationär spezifizierten Modellen herangezogen.

Bevor die Modelle miteinander verglichen werden, muss an dieser Stelle noch eine Besonderheit der Spezifikation der bedingten Wahrscheinlichkeiten in der stationär spezifizierten LTA erwähnt werden. In der Literatur waren zu gegebenem Zeitpunkt keine Referenzen oder Anmerkungen auffindbar, wie diese Spezifikation generell zu handhaben ist. Deshalb wird sie hier anschaulich dargestellt und für die weiteren Berechnungen vom Autor der vorliegenden Arbeit einheitlich fixiert.

Um die Wahrscheinlichkeiten in der Übergangswahrscheinlichkeitsmatrix zu berechnen, müssen jene Parameter, deren Notation zuvor aufgeführt wurde, wie in Tabelle 87 geschätzt werden. Diese Spezifizierung beschreiben Muthén und Muthén (2009). Weitere LTA Spezifikationen führen auch Muthén (2006, 2007) und Muthén und Muthén (2007) an. Die Parameter der Klasse drei zum Zeitpunkt zwei werden für die Schätzung auf null gesetzt. Um die Wahrscheinlichkeit eines Feldes in der Übergangsmatrix zu schätzen, müssen dafür die geschätzten Klassenzugehörigkeiten zum Zeitpunkt t und $t-1$, und die geschätzte Übergangswahrscheinlichkeit, bekannt sein. Für die Übergangsmatrizen zwischen Zeitpunkt zwei und drei, drei und vier, vier und fünf, etc. entsteht dadurch auch kein Problem. Für die Übergangsmatrix zwischen Zeitpunkt eins und zwei führt dies jedoch zu einem Spezialfall.

		C_2		
		1	2	3
C_1	1	a_1+b_{11}	a_2+b_{21}	0
	2	a_1+b_{12}	a_2+b_{22}	0
	3	a_1	a_2	0

Tabelle 87: Parameterspezifikationen im LTA Modell (Muthén und Muthén, 2009)

4.3.7.20. Modellvergleich unter Verwendung unterschiedlicher Zeitspannen - Kinodatensatz

Um das Modell stationär definieren zu können, ist die Wahrscheinlichkeit der Klassenzugehörigkeit zum Zeitpunkt eins nicht notwendig. Schon durch die Gleichsetzung der Klassenwahrscheinlichkeitszugehörigkeiten für die Zeitpunkte

zwei, drei, vier, etc. entsteht eine gleichbleibende Wahrscheinlichkeit für a_1 , also ein stationärer Prozess. Die Klassenzugehörigkeitswahrscheinlichkeit zu Klasse eins zum Zeitpunkt zwei, also a_1 , muss jedoch nicht von der Wahrscheinlichkeit der Klassenzugehörigkeitswahrscheinlichkeit der Klasse drei zum Zeitpunkt eins, also wiederum a_1 , abhängig gemacht werden. Für den stationären Prozess bedeutet dies folgendes: Wird die Klassenzugehörigkeitswahrscheinlichkeit zu Klasse drei zum Zeitpunkt eins mitgeschätzt und mit den a_1 Parametern zwischen den folgenden Zeitpunkten gleichgesetzt, so muss ein Parameter weniger geschätzt werden, was zur Einsparung eines Freiheitsgrades führt. Wird diese nicht den weiteren a_1 Parametern gleichgesetzt, so muss diese separat geschätzt werden und dies führt zum Verlust eines Freiheitsgrades. Um zu einer Entscheidung zu kommen, ob der zusätzliche Freiheitsgrad und die einhergehende Veränderung der *loglikelihood* zu signifikanten Veränderungen führen, wird hier ein LRT durchgeführt. Es wird von genesteten Modellen ausgegangen. Der Vergleich wird in Tabelle 88 anhand der stationären Modelle mit zwei Klassen gezeigt.

Modell		Log Likelihood	BIC	ABIC	Freie Parameter
Zwei Klassen	fixiert	-7.433,383	14.956,286	14.918,163	12
sechs Zeitpunkte	frei	-7.425,748	14.948,475	14.907,175	13
Zwei Klassen	fixiert	-14.672,492	29.434,504	29.396,381	12
zwölf Zeitpunkte	frei	-14.631,739	29.360,457	29.319,157	13

Tabelle 88: Spezifikationsvergleich

Im Zwei-Klassenmodell mit sechs Zeitpunkten ergibt sich eine *loglikelihood*-Differenz von 7,635 und einem Freiheitsgrad, was einen Signifikanzwert von 0,0057246 ergibt. Im Zwei-Klassenmodell mit zwölf Zeitpunkten ergibt sich eine *loglikelihood*-Differenz von 40,753 und einem Freiheitsgrad, was einen Signifikanzwert von $1,7274e^{-10}$ ergibt. Somit ist das weniger eingeschränkte Modell mit dem zusätzlich geschätzten Parameter signifikant zu bevorzugen. Kaplan (2008) setzt den Parameter in den stationären Modellen mit zwei Klassen den weiteren Parametern gleich, ohne jedoch auf die Spezifikation in der Mplus-Syntax im Anhang zu verweisen. In den Modellen mit drei Klassen von Nylund (2007) wird der Parameter nicht gleichgesetzt. Dieses Vorgehen wird in der vorliegenden Arbeit verwendet.

4.3.7.21. *Stationary vs. nonstationary latent transition analysis* - Kinodatensatz

Tabelle 89 zeigt einen Überblick der unterschiedlichen schätzbaren Modelle. Mit Hilfe des LRT dürfen nur stationäre und nicht-stationäre Modelle selbiger Klassenanzahl miteinander verglichen werden. Der Vergleich zwischen dem stationären und nicht-stationären Zwei-Klassenmodell über sechs Zeitpunkte zeigt eine Devianz von 25,663 mit neun Freiheitsgraden und ist mit einem Wert von 0,0023190 hoch signifikant. Der Vergleich zwischen dem stationären und nicht-stationären Zwei-Klassenmodell über zwölf Zeitpunkte zeigt eine Devianz von 78,645 mit 21 Freiheitsgraden und ist mit einem Wert von $1,36e^{-4}$ hoch signifikant. Der Vergleich zwischen dem stationären und nicht-stationären Drei-Klassenmodell über sechs Zeitpunkte zeigt eine Devianz von 18,147 mit 24 Freiheitsgraden und ist mit einem Wert von 0,79582 nicht signifikant. Letzterer ist jedoch mit Vorsicht zu genießen, da aufgrund von Problemen der Nichtidentifizierung des nicht-stationären Modells einige Parameter im Zuge der Schätzung fixiert werden mussten.

	Stationär			Nicht stationär		
	6 Zeitpunkte		12 Zeitpunkte	6 Zeitpunkte		12 Zeitpunkte
	2 Klassen	3 Klassen	2 Klassen	2 Klassen	3 Klassen	2 Klassen
Log Likelihood	-7.433,383	-7.320,269	14.672,492	-7.407,720	(-7.302,122)	14.593,847
BIC	14.956,286	14.812,117	29.434,504	14.972,099	(14.954,860)	29.433,872
ABIC	14.918,163	14.739,048	29.396,381	14.905,384	(14.805,545)	29.329,034
Freie Parameter	12	23	12	21	47 (Warnung)	33

Tabelle 89: Modellvergleich

4.3.7.22. *Modellvergleich mit unterschiedlicher Klassenanzahl im Messmodell* - Kinodatensatz

Für den Vergleich zwischen den Modellen mit unterschiedlicher Klassenanzahl kann BIC herangezogen werden, LRT ist nicht geeignet. Jedoch kann der Unterschied der *loglikelihood* und der Unterschied der Anzahl an freigeschätzten Parametern verwendet werden, um zumindest einen Eindruck zur Modellverbesserung zu gewinnen. Der Vergleich zwischen dem stationären Zwei- und Drei-Klassenmodell über sechs Zeitpunkte zeigt einen BIC-Unterschied von 144,169 und eine *loglikelihood*-Differenz von 113,114 bei elf Freiheitsgraden und deutet auf eine Verbesserung durch Hinzunahme einer

weiteren Klasse hin. Der Vergleich zwischen dem nicht-stationären Zwei- und Drei-Klassenmodell über sechs Zeitpunkte zeigt einen BIC-Unterschied von 17,239 und eine *loglikelihood*-Differenz von 105,598 bei 26 Freiheitsgraden. Zumindest die *loglikelihood*-Differenz deutet ebenfalls auf eine Verbesserung durch Hinzunahme einer weiteren Klasse hin. Letzterer Vergleich ist aus bereits genannten Gründen mit Vorsicht zu genießen.

4.3.7.23. *Non-stationary two-class model vs. stationary three-class model - Kinodatensatz*

Da das nicht-stationäre Drei-Klassenmodell Probleme bei der Schätzung aufweist, wird für die weitere Auswahl das nicht-stationäre Zwei-Klassenmodell mit dem stationären Drei-Klassenmodell verglichen. Im Drei-Klassenmodell ist das am häufigsten auftretende Pattern jenes, in der gleichen Klasse über die Zeit hinweg zu verbleiben. Dies ist Klasse eins mit dem Pattern 1 1 1 1 1 mit 1.464 Probanden, laut der *most likely latent class membership*, von insgesamt 1.737 Probanden. Die Klassenzugehörigkeitsveränderung ist aus Tabelle 90 ersichtlich.

	1	2	3
1	0,962	0,025	0,012
2	0,632	0,129	0,239
3	0,005	0,057	0,939

Tabelle 90: Sechs Zeitpunkte - drei Klassen - stationär - invariantes Messmodell

Es ist eindeutig erkennbar, dass die Klasse eins und drei am stabilsten über die Zeit hinweg sind. Die Personen aus Klasse zwei tendieren dazu, über die Zeit in die Klasse eins zu hüpfen. Die Verteilungen der Klassen über die Zeit aufgrund ihrer *most likely latent class membership* sind aus Tabelle 91 ersichtlich.

Diagramm 136 blendet die Wahrscheinlichkeiten mit ‚sehr zufrieden (1)‘ zu antworten für die einzelnen Klassen auf. Diese Wahrscheinlichkeiten bleiben über die Zeit hinweg gleich, da das Messmodell über die sechs Zeitpunkte konstant gehalten wird. Es ist erkennbar, dass die drastischen Unterschiede wie im LCA-Modell verschwinden und sich eine sehr zufriedene, eine mittel zufriedene und eine weniger zufriedene Gruppe in allen fünf Zufriedenheitsitems herauskristallisiert.

Zeitpunkt	Klasse	Absolute Anzahl	Prozentueller Anteil
1	1	1.464	84,283
	2	35	2,015
	3	238	13,702
2	1	1.462	84,168
	2	34	1,957
	3	241	13,874
3	1	1.450	83,477
	2	45	2,591
	3	242	13,932
4	1	1.471	84,686
	2	23	1,324
	3	243	13,990
5	1	1.469	84,571
	2	29	1,670
	3	239	13,759
6	1	1.467	84,456
	2	33	1,900
	3	237	13,644

Tabelle 91: Respondentenverteilung über die Klassen

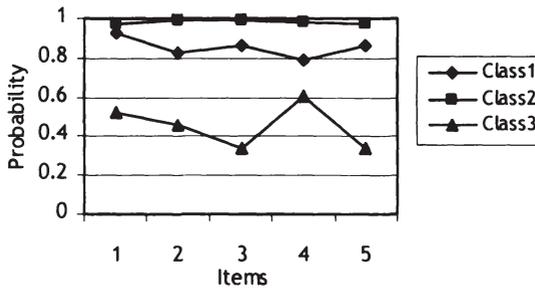


Diagramm 136: Item probability plot

4.3.7.24. *Non-stationary two-class model mit full measurement invariance - Kinodatensatz*

Nun wird mit der nicht-stationären Zwei-Klassenlösung mit invariantem Messmodell verglichen. Die Klassenzugehörigkeitsveränderung in Tabelle 92 ist zu jedem Zeitpunkt unterschiedlich.

		Zeitpunkt 2	
		Klasse 1	Klasse 2
Zeitpunkt 1	Klasse 1	0,904	0,096
	Klasse 2	0,562	0,438
		Zeitpunkt 3	
		Klasse 1	Klasse 2
Zeitpunkt 2	Klasse 1	0,892	0,108
	Klasse 2	0,241	0,759
		Zeitpunkt 4	
		Klasse 1	Klasse 2
Zeitpunkt 3	Klasse 1	0,967	0,033
	Klasse 2	0,413	0,587
		Zeitpunkt 5	
		Klasse 1	Klasse 2
Zeitpunkt 4	Klasse 1	0,948	0,052
	Klasse 2	0,273	0,727
		Zeitpunkt 6	
		Klasse 1	Klasse 2
Zeitpunkt 5	Klasse 1	0,948	0,052
	Klasse 2	0,236	0,764

Tabelle 92: Sechs Zeitpunkte - zwei Klassen - nicht-stationär - invariables Messmodell

Es ist erkennbar, dass die Wahrscheinlichkeit zu den verschiedenen Zeitpunkten in Klasse eins zu verbleiben sehr viel größer ist, verglichen mit den Wechselwahrscheinlichkeiten und Wahrscheinlichkeiten der Klasse zwei. Die Verteilungen der Klassen über die Zeit aufgrund ihrer *most likely latent class membership* sind in Tabelle 93 aufgelistet.

Diagramm 137 blendet die Wahrscheinlichkeiten mit ‚sehr zufrieden (1)‘ zu antworten für die beiden Klassen auf. Diese bleiben über die Zeit aufgrund des konstanten Messmodells ebenfalls gleich. Auch die *cross-sectional-LCA* Modellunterschiede verschwinden und es kristallisiert sich eine sehr und eine mittelmäßig zufriedene Gruppe heraus.

Zeitpunkt	Klasse	Absolute Anzahl	Anteil
1	1	1.608	92,573%
	2	129	7,427%
2	1	1.575	90,674%
	2	162	9,326%
3	1	1.487	85,607%
	2	250	14,393%
4	1	1.553	89,407%
	2	184	10,593%
5	1	1.554	89,465%
	2	183	10,535%
6	1	1.540	88,659%
	2	197	11,341%

Tabelle 93: Respondentenverteilung über die Klassen

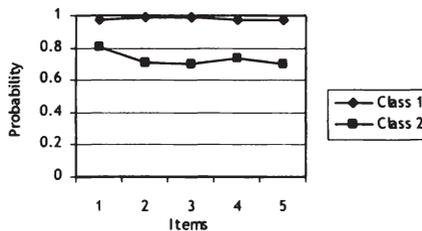


Diagramm 137: Item probability plot

Zusammenfassend kann festgestellt werden, dass die Unterschiede in den Lösungen zur inhaltlich gleichen Aufspaltung der Gruppen führen, nämlich nach dem Grad der Wahrscheinlichkeit einer positiven Zufriedenheitsantwort. Verglichen mit der Zwei-Klassenlösung weist die Drei-Klassenlösung einen zusätzlichen Cluster auf, welcher Probanden beinhaltet, für welche die Wahrscheinlichkeit einer positiven Zufriedenheitsantwort sehr gering ist. Diagramm 138 verdeutlicht den Unterschied. Die größte sehr zufriedene Klasse ist in beiden Gruppen nahezu ident und deren Profile überlappen sich. So liegt die inhaltliche Aufspaltung der weniger zufriedenen Klasse bei der Zwei-Klassenlösung, zwischen den beiden weniger zufriedenen Clustern der Drei-Klassenlösung.

Erweiterungen sind beispielsweise über die Einführung von Kovariaten möglich. Die folgenden beiden Artikel liefern Beispiele dafür, ersterer die Hintergründe zur Einführung, letzterer ein anspruchsvolles Anwendungsbeispiel.

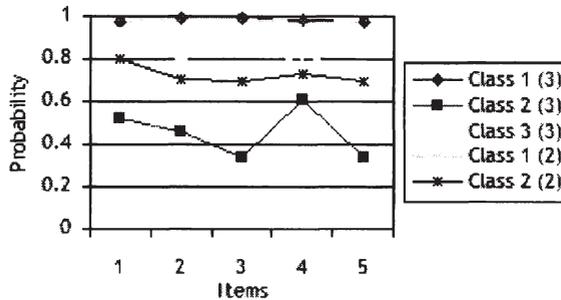


Diagramm 138: Item probability plot Vergleich

4.3.7.25. Kovariaterweiterungen

Vermunt et al. (1999) stellen ein *discrete-time discrete-state latent Markov (LM)* Modell mit *time-constant* und *time-varying* Kovariaten zur Analyse kategorialer Variablen vor. In *discrete-time discrete-state* Markov-Modellen unterliegen die beobachteten Zustände meist Messfehlern und somit sind die Übergänge zwischen Zuständen teilweise verfälscht. LM-Modelle bieten die Möglichkeit die wahre Veränderung vom Messfehler zu trennen. Das *standard* LM-Modell ist aber begrenzt, wenn es darum geht, individuelle Unterschiede in der Wahrscheinlichkeit einen Zustand zu einem Zeitpunkt anzunehmen, zu messen. Deshalb stellen sie ein flexibles *logit*-Regressionsmodell vor, welches erlaubt, die latenten Zustände zu den jeweiligen Zeitpunkten auf *time-constant* und *time-varying* Kovariate zu regressieren. Somit werden *causal log-linear*-Modelle mit LC-Modellen mit erklärenden Variablen verbunden. Der Ansatz wird anhand des Interesses an Physik (A) gezeigt, welches durch die *time-constant* Kovariate Geschlecht (X) und die *time-varying* Kovariate Schwierigkeitsstufe der Physikklasse (Z) erklärt wird. W beschreiben die latenten Zustände in Diagramm 140.

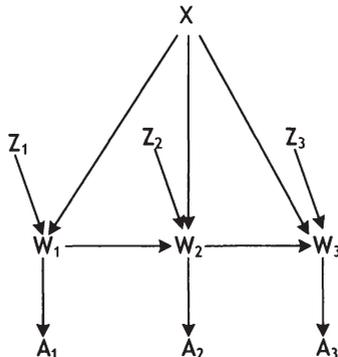


Diagramm 139: Discrete latent Markov Modell mit Kovariaten (Vermunt et al., 1999)

Sind die beobachteten Zustände fehlerbehaftet, so sind die beobachteten Übergänge eine Mischung aus wahrer und verfälschter Veränderung, welche durch die Messfehler in den beobachteten Zuständen ausgelöst wird. LM-Modelle trennen wahre Veränderung und Messfehler. Das *standard-LM* Modell ist begrenzt, wenn es darum geht, individuelle Unterschiede in der Wahrscheinlichkeit einen Zustand zu einem Zeitpunkt anzunehmen, zu messen. Die einzige Möglichkeit die beobachtete Heterogenität zu berücksichtigen ist die *multiple-group-Analyse*. Ein Nachteil ist die hohe Anzahl an Parametern, wenn mehrere erklärende Variable in die Analyse eingebracht werden. Weiters funktioniert sie nur für *time-constant* Kovariate. Es wird also ein Modell benötigt, welches den latenten Zuständen erlaubt, sowohl *time-constant* als auch *time-varying* Kovariate zu inkludieren. *Goodman's causal log-linear* Modell ist ein rekursives System aus *logit*-Modellen, in welchen eine Variable, welche in einer Gleichung als abhängige Variable auftaucht, als unabhängige Variable in einer nachfolgenden Gleichung auftauchen kann. Vermunt et al. (1999) präsentieren ein LM-Modell, in welchem latente Zustände auf *time-constant* und *time-varying* Kovariate durch ein System an *logit*-Modellen regressiert werden. Dies ist eine Erweiterung von *Goodman's causal log-linear*-Modellen.

Manifeste Markov-Modelle betrachten wiederholte Beobachtung kategorialer oder diskreter Variablen. Ist die Variable diskret und die Messung erfolgt zu mehreren Zeitpunkte, wird das Modell *discrete-time discrete-space* Markov-Modell genannt. Es wird angenommen, dass die Variable ohne Fehler gemessen wird. Im LC-Modell werden Messfehler berücksichtigt. Auf Messfehler wird in der Familie der *latent structure*-Modelle bedacht, welche alle der Annahme lokaler Unabhängigkeit folgen. Das bedeutet, dass die beobachteten Variablen oder Indikatoren, welche zur Messung der nicht beobachteten Variable dienen, gemeinsam unabhängig für einen bestimmten Wert der nicht beobachteten oder latenten Variablen sind. In der Faktorenanalyse werden kontinuierliche manifeste Variablen als Indikatoren einer oder mehrerer kontinuierlicher latenter Variablen verwendet. In *latent trait*-Modellen wird eine kontinuierliche Variable als zugrundeliegend für ein Set an kategorialen Indikatoren verwendet. Sind sowohl die manifesten als auch die latenten Variablen kategorial, ergibt dies das LC-Modell. Im LM-Modell werden Markov- und LC-Modell verbunden, um Zustandsveränderungen zu unterschiedlichen Zeitpunkten zu analysieren, wobei diese mit Fehlern gemessen wurden. Die Beziehung zwischen zwei beobachteten Variablen mit Messfehlern wird für gewöhnlich schwächer sein, als die wahre Beziehung. Die Beziehung wird also unterschätzt bzw. die Höhe der Veränderung überschätzt. Unterliegen die Daten den Messfehlern, so sind die beobachteten Übergänge eine Mischung aus wahrer und verfälschter

Veränderung. LM-Modelle separieren wahre und verfälschte Veränderung, welche durch den Messfehler erzeugt wird. Im Vergleich zum LC-Modell ist es möglich ein LM-Modell mit nur einem Indikator zu jedem Beobachtungszeitpunkt zu schätzen.

Es kann nicht davon ausgegangen werden, dass der betrachtete Prozess für alle Personen einer Population gleich ist. So existieren für LCA-Erweiterungen für Subpopulationen, oder auch *multiple-group*-LM-Modelle. Diese Markov-Modelle inkludieren eine zusätzliche Variable, welche die Subgruppenzugehörigkeit beschreibt. Jeder Parameter wird als subgruppenspezifisch angenommen. Es ist möglich das Modell durch Parametergleichsetzung über Subgruppen zu beschränken. So wird meist der Messfehler über die Gruppen gleichgesetzt. Auch die anfängliche Verteilung oder Übergangswahrscheinlichkeiten können für alle Gruppen gleichgesetzt werden. Dabei gibt es aber Beschränkungen. In LM-Modellen ist die Bestimmung der Kovariateneffekte sowohl für die anfängliche Position, als auch für die Übergangswahrscheinlichkeiten von Interesse. In der *multiple-group*-Analyse können nur Kreuztabellen über die Kovariate gemacht und diese gemeinsame Kovariate als Gruppierungsvariable verwendet werden. Dies macht nur Sinn, wenn die Zellenanzahl der gemeinsamen Verteilung der unabhängigen Variablen nicht zu groß ist, da andererseits eine große Parameteranzahl geschätzt werden muss. Weiters kann der Einfluss auf die Zustände zu unterschiedlichen Zeitpunkten *time-varying* sein. In der *multiple-group* Analyse ist dies sehr schwierig zu gebrauchen. Somit ist ein regressionsähnliches Modell nötig, welches sowohl *time-constant* als auch *time-varying* Kovariate verarbeiten kann. Das vorgestellte *logit* Regressionsmodell, speziell ein *causal log-linear*-Modell, wird zuerst erweitert, um den Nachteil des *multiple-group*-Modells zu umgehen, nämlich dass die Gruppierungsvariable *time-constant* sein muss. Nachfolgend wird durch Dimensionsreduktion der marginalen Tabellen eine effizientere Berechnung geschaffen. Im *Goodman's causal log-linear* Modell wird angenommen, dass alle Variablen beobachtet wurden. Es ist jedoch erwünscht, latente Zustände auf vorangehende latente Zustände, *time-constant* und *time-varying* Kovariate zu regressieren. In den *causal log-linear*-Modellen mit latenten Variablen wird ein Regressionsmodellansatz erschaffen, welcher es erlaubt, eine latente Variable auf ein Set von Kovariaten zu beziehen und wie in den *causal log-linear*-Modellen, sowohl der abhängigen Variable, als auch der Kovariaten, sich über die Zeit hinweg zu verändern. Es wird ein simultanes System an *logit*-Gleichungen, für ein Set aus kausal geordneten latenten und manifesten Variablen und ein LC-Modell für latente Variablen, welche in den *logit*-Modellen verwendet werden, geschaffen.

Ein äußerst interessanter Punkt ist auch die Begründung, warum die Physikschwierigkeitsstufen in den Schulen als *time-varying* Kovariate eingebracht werden. Ausgehend von der im Theorieteil bereits erwähnten Motivations-Theorie, wird das Interesse bzw. die intrinsische Motivation durch das Kompetenzgefühl beeinflusst, welches in den höher eingestuft Klassen besser sein sollte. Das Interesse an Physik zu einem bestimmten Zeitpunkt (W) hängt vom Physikinteresse zum vorangehenden Zeitpunkt (W), vom Geschlecht (X) und von der Schwierigkeitsstufe der Physikklasse zum gleichen Zeitpunkt (Z) ab, vorausgesetzt es gibt nur einen Zwei-Variablen-Effekt. Dies beinhaltet die Markov-Annahme, bei welcher es keine zeitverzögerten Effekte gibt und der Effekt von Geschlecht und dem Schulschwierigkeitsgrad unabhängig vom vorangegangenen Interesse sind. Weiters muss ein gleichbleibender Messfehler angenommen werden, um das *single indicator*-LM-Modell identifizierbar zu machen. Das Modell im vorliegenden Fall passt nicht und es muss zumindest eine Annahme in den jeweiligen Modellen aufgeweicht werden. Im ersten Modell wird ein direkter Effekt von W_1 auf W_3 angenommen, wodurch die Markov-Annahme nicht gehalten werden kann. Im Beispiel wurden die Markov-Annahme und die Annahme, dass Z_t nicht durch W_{t-1} beeinflusst wird, abgelehnt. Weiters werden Beschränkungen hinsichtlich der Effekte über die Zeit eingeführt, um unnötige Parameter aus dem Modell zu nehmen. So werden die Effekte, wie in Diagramm 141 dargestellt, von W_{t-1} auf W_t , von X auf W_t , von Z_t auf W_t und von W_t auf Z_{t+1} als zeitunabhängig angenommen, also zeitliche Homogenitätsrestriktionen.

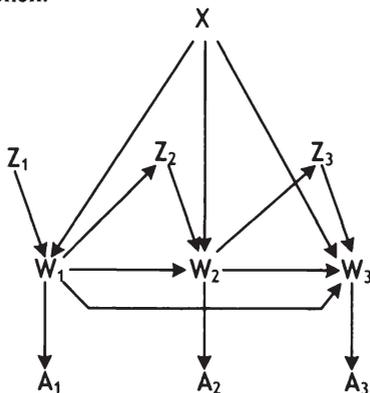


Diagramm 140: Discrete latent Markov Modell mit Kovariaten und weiteren Effekten (Vermunt et al., 1999)

Um zu testen ob der Messfehler vernachlässigbar ist, wird von einer perfekten Messung der Antwortwahrscheinlichkeiten ausgegangen und diese werden

fixiert. Es ist zu beachten, dass selbiges Modell ohne Beschränkung, eben die Fixierung der Parameter auf gleiche Randwerte, nicht mehr mit dem Modell mit Beschränkung über LRT vergleichbar ist. Die *time-varying* Kovariate senkt den Wert des Messfehlers. Die Gründe dafür liegen möglicherweise darin, dass die nicht beobachtete Heterogenität in den Zuständen durch den Messfehler aufgefangen wurde, was durch die Inkludierung von Z verschwunden ist. Daher ist in LM-Ketten mit einem einzelnen Indikator pro Beobachtung die Unterscheidung zwischen Messfehler und nicht beobachteter Heterogenität schwierig. Um Messfehler aufzudecken, sind mehr Indikatoren pro Beobachtungszeitpunkt empfehlenswert, da in diesem Fall die Beziehung zwischen den Indikatoren Information über die Reliabilität jedes Indikators bringt.

Zusammengefasst konnten Vermunt et al. (1999) zeigen, dass die *first-order*-Markov-Annahme für das Interesse an Physik nicht gehalten werden kann, es *time-homogeneous* Effekte der *time-constant* Kovariaten Geschlecht und *time-varying* Kovariate Physikinteressensgrad gibt, und eine indirekte Beziehung zwischen Physikinteresse nachfolgender Zeitpunkte und dem Schwierigkeitsgrad der Physikkategorie existiert. Der geschätzte Wert des Messfehlers in den beobachteten Zuständen ist vernachlässigbar. Da jedoch nicht davon ausgegangen werden kann, dass das Physikinteresse ohne Fehler gemessen wurde, wird dies als Resultat, dass nur ein Indikator pro Beobachtung verwendet wurde, angesehen. In vorgestelltem LM-Modell wurde angenommen, dass nur die abhängige Variable Messfehler unterliegt. Dies kann jedoch auch für Kovariate erweitert werden. Auch kontinuierliche *time-constant* Kovariate können hinzugenommen werden, was jedoch die Testung des Modellfits über die χ^2 -*statistic* nicht mehr erlaubt. Abschließend erwähnen sie noch eine Einschränkung. Es ist ein generelles Problem der Analyse kategorialer Daten, dass bei spärlichen Tabellen die theoretische χ^2 -Annäherung der Pearson- χ^2 -*statistic* und der LR- χ^2 -*statistic* schwach ist. Obwohl in solchen Situationen die Parametersignifikanz mit Hilfe des *conditional* LRT getestet werden kann, kann der Fit des Modells nicht mehr bestimmt werden. Eine mögliche Lösung bietet die *bootstrap*-Prozedur zur Modelltestung.

Chung et al. (2005) liefern ein Beispiel einer LTA mit Kovariaten. Sie messen mit einer erweiterten LTA mit zwei latenten Variablen, nämlich dem pubertären Status und der sequentiellen Substanzgebrauchszustände, den Einfluss der pubertären Entwicklung, des Alters und der Interaktion auf den jugendlichen Gebrauch von Substanzen. Um eine Serie an Zuständen zu identifizieren, modellieren sie einen stufenweisen Prozess. Der Gebrauch von Substanzen ist

time-varying. Die Altersabhängigkeit wird als Kovariate in Form einer logistischen Regression eingebracht. Basierend auf einem Fünf-Zustände-Modell, zeigen drei Klassen unterschiedliche Patterns an Zuständen was den Substanzgebrauch betrifft. Mit Hilfe polytomer diskreter Variablen wird der Substanzgebrauch, *stage item*, und mit diskreten Variablen die pubertäre individuelle Entwicklung, *class item*, beschrieben. *Stage-items* stehen für den Zustand des Substanzgebrauchs, die *latent stage*-Variable. *Class-items* stehen für die Klasse der pubertären Entwicklung, *latent class*-Variable. Ein Vektor von Kovariaten kann als diskret bzw. kontinuierlich und auch zeitabhängig ins Modell eingebracht werden. Das Modell wird als *first-order* Markovkette spezifiziert. Die Itemantwortwahrscheinlichkeiten repräsentieren die Wahrscheinlichkeit einer Antwort auf ein *class-item*, gegeben die pubertäre Klassenzugehörigkeit. Weiters wird die Wahrscheinlichkeit einer Antwort auf ein *stage-item*, gegeben die Zugehörigkeit zu einem Substanzgebrauchszustand zu einem gewissen Zeitpunkt und Klassenzugehörigkeit geschätzt. Es gibt zwei binär codierte *class-items*, um die weibliche physische Reife zu messen, nämlich die Brustgröße verglichen mit der Schulstufe und die Ausprägung der Kurven des Körpers verglichen mit der Schulstufe. Dadurch werden drei Klassen der pubertären Entwicklung bestimmt. Für die *stage-items* wird der jugendliche Substanzgebrauch anhand der Items, Alkoholkonsum in den letzten zwölf Monaten, Zigarettenkonsum in den letzten 30 Tagen, fünf oder mehr alkoholische Getränke auf einmal in den letzten zwölf Monaten und Trunkenheit in den letzten zwölf Monaten, gemessen. Die Exponenten der Regressionskoeffizienten werden als geschätzte *odds ratios* interpretiert. Graphiken der marginalen Verbreitung der Substanzgebrauchszustände über das Alter und die relativen Anteile über das Alter werden zu Interpretationszwecken herangezogen. Die Übergangswahrscheinlichkeiten zwischen den Substanzgebrauchszuständen zwischen den Zeitpunkten werden für jede pubertäre Klasse separat errechnet.

Chung et al. (2005) erwähnen Pearson- χ^2 und die LR-*statistic* G^2 als Standardstatistiken um den Fit einer LTA zu messen, wobei der Fit des geschätzten Modells mit dem saturierten Modell verglichen wird. Jede der Statistiken ist asymptotisch χ^2 -verteilt mit Freiheitsgraden, nämlich die Anzahl der möglichen Antwortpatterns minus die Anzahl an geschätzten Parametern minus eins. In LTA-Modellen kann die asymptotische Annahme nicht gehalten werden, da diese große Kontingenztafeln mit großer Anzahl an Freiheitsgraden mit sich bringen. Deshalb stellen sie eine graphische Lösung für die Beurteilung des Fits des gesamten Modells vor, wobei sie das geschätzte LTA-Modell mit den empirischen Daten vergleichen. Es werden die standardisierten Residuen graphisch dargestellt. Die standardisierten Residuen für die drei unter-

schiedlichen Altersgruppen und zwei Wellen werden geplottet. Zwei LTA-Modelle, erstes mit dem Alter als Kovariate und zweites ohne, werden zur Bestimmung der Verbesserung des Modells durch Miteinbeziehung des Alters als Kovariate verglichen. Die standardisierten Residuen werden durch $r_j = (\hat{f}_j^{SAT} - \hat{f}_j^{LT}) / \sqrt{\hat{f}_j^{LT}}$ berechnet, wobei \hat{f}_j^{SAT} bzw. \hat{f}_j^{LT} die erwarteten Häufigkeiten für jeweils ein Antwortpattern des saturierten Modells bzw. eines der beiden LTA-Modelle, darstellen. Das LTA-Modell mit dem Alter passt besser zu den Daten als das Modell ohne dessen Miteinbeziehung.

Erweiterungen über Kovariate werden im Kinobeispiel nicht eingeführt um die Komplexität geringer zu halten. Jedoch auch theoretisch begründete Erweiterungen hinsichtlich der Spezifikation der Übergangswahrscheinlichkeitsmatrix sind möglich. Die folgenden drei Artikel beschreiben Entwicklungen spezieller Restriktionsmöglichkeiten.

4.3.7.26. *Longitudinal Guttman Simplex*

Collins et al. (1988) stellen ein Modell für dynamische Konstrukte, nämlich den *longitudinal Guttman simplex* (LGS) vor. Dieser ist ein Spezialfall des *Guttman simplex*. Es wird zusätzlich zu Personen und Items, die Zeit miteinbezogen. Zugrunde liegt dem LGS die gemeinsame Ordnung. Personen erschaffen eine Schwierigkeitsordnung von Items und Items eine Fähigkeitsordnung für Personen. Diese Konsistenz erlaubt eine gemeinsame Ordnung von Personen und Items in die Guttman-Skala. Im LGS existiert für jede Person eine gemeinsame Item-Zeit-Ordnung welche über die Personen konsistent ist. Die Antworten einer Person auf ein Set von Items zu unterschiedlichen Zeitpunkten ermöglichen eine Schwierigkeitsordnung der Items und eine zeitliche Ordnung der Zeitpunkte. Die Item-Zeit-Ordnung ist auch über die Personen konsistent. Somit ist die gemeinsame Ordnung von Personen über Zeitpunkte für jedes Item konsistent und auch die gemeinsame Ordnung von Personen und Items für jeden Zeitpunkt. Es wird eine kumulative einheitliche Entwicklung unterstellt. Somit durchläuft jede Person Entwicklungsstadien einer Guttman-Skala über die Zeit und zwar in dieselbe Richtung. Damit diese auch kumulativ ist, müssen die Fähigkeiten über die Zeit erhalten bleiben. Einheitlich ist die Entwicklung dann, wenn jede Person sich nur in die gleiche Richtung entwickeln kann. Entscheidend ist auch die Möglichkeit der Entwicklung zwischen unterschiedlichen Stufen zu unterschiedlichen Raten und die Brauchbarkeit dieser Modellklasse für psychosoziale Konstrukte wird explizit erwähnt. In Tabelle 94 werden die relevanten Antwortabfolgen des LGS dargestellt. Sie benötigen diese für die Entwicklung eines Konsistenzindex, wobei diese noch gewichtet werden, da sie unterschiedlich starke Einflüsse auf das Vorliegen einer longitudinalen Guttman-

Skala haben. *Redundant time relation* geben Auskunft über die Schwierigkeitsunterschiede zu unterschiedlichen Zeitpunkten. *Redundant item relations* geben Auskunft über die Schwierigkeitsunterschiede zwischen Items. *Unique relations* sind von speziellem Interesse, da diese die gemeinsame Ordnung von Items und Zeitpunkten reflektieren. *Contradictory relations* stellen inkonsistente Ordnungen zwischen Items und Zeitpunkten dar.

Relations	Redundant time		Redundant item		Unique		Contradictory	
	Item A	Item B	Item A	Item B	Item A	Item B	Item A	Item B
Time 1	fail	fail	pass	fail	pass	pass	fail	pass
Time 2	pass	pass	pass	fail	fail	pass	pass	fail

Tabelle 94: Longitudinal Guttman Simplex Antwortverhalten (Collins et al., 1998)

Inkonsistenzgründe der vorgeschlagenen Skala liegen in der schlechten Datenqualität, an Items, welche unempfindlich auf Veränderungen reagieren, oder am zugrundeliegenden Konstrukt, welches nicht einer kumulativen einheitlichen Entwicklung unterliegt. Collins et al. (1988) erwähnen auch die Brauchbarkeit der Modelle aufgrund der Möglichkeit, Stufen zu überspringen, Rückschritte zu vorangegangenen Stufen oder vielfache Entwicklungen. Für das Kinobeispiel wird die Annahme eines LGS aufgrund der Anpassung der Servicequalitätslevels seitens der Kinobetreiber und auch der Ansprüche der Kunden nur schwer zu halten sein. Einzig die vielfach postulierte Anpassung an die Servicequalitätslevels mit einer einhergehenden konstanten sinkenden Zufriedenheit stellt eine logische Übersetzung eines derartigen Modells dar.

Collins und Cliff (1990) stellen ein Beispiel für zuvor genannte Methode vor und besprechen die Sinnhaftigkeit in Bezug auf Validität und Reliabilität. Die Konstruktvalidität wird durch die Einbringung der latenten Variablen in das LGS herbeigeführt. Dabei geht es um die interindividuellen Entwicklungsunterschiede. Diese können linear als auch nicht linear sein, müssen jedoch monoton sein. Sie merken an, dass traditionelle Messtheorien wie die klassische Testtheorie und die *item response theory* auf einem sich nicht verändernden wahren Wert basieren und somit für Wachstumsentwicklungen nicht brauchbar sind. Die Messung des monotonen individuellen Wachstums wird durch eine genauere Auswahl an Items zwecks Entwicklung eines Messinstrumentes verbessert. Traditionelle Ansätze basieren nicht auf der Entwicklung individuellen Wachstums einer latenten Variable und somit wird die Konstruktvalidität schlechter. Weiters können diese die intraindividuelle Variabilität über die Zeit nicht einfangen. Items, welche keine *within-time* interindividuelle Variabilität aufweisen, reduzieren in herkömmlichen Modellen

die Skalenkonsistenz und werden aus den Modellen entfernt, da eine Population mit einem hohen *within-time* interindividuellen wahren Varianzwert eine höhere Reliabilität erhält, verglichen mit Populationen mit einem geringen *within-time* interindividuellen wahren Varianzwert. Somit kann die Messung einer homogenen Population keine hohe Reliabilität aufweisen, auch wenn diese ein präzises Wachstum aufweist. Geben alle Personen zu den unterschiedlichen Zeitpunkten die gleiche Antwort, kann die Schwierigkeit der Items nicht festgestellt werden. Werden die Antworten jedoch zu den unterschiedlichen Zeitpunkten gesamt betrachtet, kann durch Identifizierung einer longitudinalen Guttman-Entwicklung diese identifiziert werden. Erst die Sequenz über die Zeitpunkte lässt Rückschlüsse auf die Brauchbarkeit eines Items zu. Es geht also um die Item-Zeitpunkt-Ordnung. Items können zu einem Zeitpunkt eine hohe Varianz aufweisen und somit für nur einen Zeitpunkt betrachtet tauglich sein. Über mehrere Zeitpunkte kann sich jedoch herausstellen, dass kein Wachstum besteht und das Item zur Messung einer Entwicklung nicht tauglich ist, trotz dessen, dass herkömmliche Methoden das Item als tauglich in eine Schwierigkeitsstufe einstufen würden. Eine Population, deren Individuen das selbige Wachstum zeigen, führen weiters zu höheren *test-retest*-Korrelationen, verglichen mit einer Population, in welcher die Individuen unterschiedliches Wachstum aufweisen. Sie postulieren, dass ein Evaluierungsmaß für diese Zwecke auf intraindividuellen Unterschieden, also individuellen Wachstumsverläufen basieren muss und nicht auf interindividuellen Unterschieden. Dafür wird CL, ein Konsistenzindex für longitudinale Guttman-Skalen entwickelt, welcher auf Ordnungsrelationen basiert. Jene die zu einer kumulativen unitären Entwicklung passen verbessern den Index, Verstöße verschlechtern ihn, siehe vorangegangene Tabelle. Vorteile liegen in der Nichtbestrafung unterschiedlicher Entwicklungsverläufe, er ist für lineare als auch unterschiedliche nicht-lineare monotone Wachstumsverläufe tauglich und kann durch die longitudinale Berücksichtigung zur Messmodellentwicklung beitragen. Die Modelle gehen jedoch von Antwortpatterns aus, welche konsistent zu einer perfekten Guttman-Skala zu jedem Zeitpunkt und auch über die Zeitpunkte hinweg sind. Auch Collins und Cliff (1990) merken an, dass die Entwicklung speziellerer Modelle anstatt des allgemeinen monotonen wünschenswert ist. Für den Kinodatensatz ist dieses Modell aufgrund bereits oben gewonnener Erkenntnisse in dieser strengen Ausführung nicht tauglich.

Auch ein absorbierender Status in einer Markovkette kann spezifiziert werden, welcher es einem Individuum nicht erlaubt, zurück zu wechseln. So erhält das Individuum eine Wechselwahrscheinlichkeit von null. Nylund (2007) beschreibt die Spezifizierung. Der Spezialfall des *longitudinal Guttman Simplex* geht von

einer monotonen Veränderung und kumulativen Wachstumsstadien aus. Sie liefert ein empirisches Beispiel für diesen Spezialfall anhand der Aneignung von Mathematikwissen, wobei immer nur genau in das nachfolgende Stadium gesprungen werden kann. Dabei wird die kategoriale latente Variable zu einem Indikator. Dieser identifiziert, welches Pattern der beobachteten Variablen erfahren wurde. Es existiert nur eine limitierte Anzahl an möglichen Patterns. Dieses Modell ist nicht sinnvoll, wenn die Veränderung nicht monoton ist, mehrere Indikatoren zur Identifizierung jedes Stadiums verwendet werden, oder wenn die Stadien keine Kombinationen binärer Indikatoren und somit nicht vorzeitig bekannt sind. Tabelle 95 weist mit ‚*‘ die frei geschätzten und mit ‚0‘ die auf null fixierten Parameter der Übergangswahrscheinlichkeitsmatrix aus.

		Grade 4			
		Addition	Subtraction	Multiplication	Division
Grade 3	Addition	*	*	0	0
	Subtraction	0	*	*	0
	Multiplication	0	0	*	*
	Division	0	0	0	*

Tabelle 95: Longitudinal Guttman Simplex (Nylund, 2007)

Eine Einschränkung der bisher anhand des Kinodatensatzes empirisch überprüften Markov-Modelle liegt in der Annahme, dass ein einzelnes Modell für die Population, aus welcher die Stichprobe gezogen wurde, als gültig angenommen wird. Es stellt sich jedoch die Frage, ob unterschiedliche Subpopulationen mit unterschiedlichen Übergangswahrscheinlichkeiten existieren. In den folgenden Ausführungen wird die Entwicklung und Spezifizierung der notwendigen Erweiterung zu *mixed* Markov-Modellen anhand der chronologischen Entwicklung vorgestellt. Es werden ausgewählte Artikel absatzweise besprochen und die methodisch vorgestellten Hintergründe kurz erörtert.

Langeheine (1988) stellt manifeste und latente Markovketten-Modelle für kategoriale Paneldaten vor. Die Nachteile der manifesten diskreten zeitlichen Markovketten-Modelle werden beschrieben und die Entwicklungen der latenten diskreten Markovketten-Modelle genutzt, welche gleichzeitig die Probleme der Heterogenität, Messfehler, Stationarität und *order*-(Ordnungs-)Effekte lösen. Manifeste stationäre Markovketten-Modelle erster Ordnung unterliegen Problemen, welche ursprünglich der Markovianischen Hypothese, also der Ordnung des Prozesses bzw. der Hypothese der zeitlichen Homogenität

(Stationarität) zugeschrieben werden. Eine Lösung zum ersten Problem der nicht passenden Markovianischen Hypothese ist die Teilung der Stichprobe in zwei Gruppen anhand einer externen Variablen. Diese fehlen jedoch häufig bzw. liefern kein zufriedenstellendes Ergebnis. Zweites wird durch ein kontinuierliches zeitliches Modell, anstatt des diskreten Zeitmodells, gelöst. Es wird folglich nach Lösungen für das Nichtpassen des manifesten Markov-Modells Ausschau gehalten.

4.3.7.27. *Mixed Markov model*

Das manifeste stationäre Markov-Modell erster Ordnung bringt einige Bedingungen mit sich. Jede Beobachtung hängt nur von der vorangehenden Beobachtung ab, oder für höhere Ordnungen von mehreren zurückliegenden Beobachtungen. Die Stationarität der Übergangswahrscheinlichkeiten, also ein Gleichbleiben der Übergangswahrscheinlichkeitsmatrizen, muss über die Zeit gewährleistet sein. Die latenten stationären Markov-Modelle erster Ordnung werden in *no latent change*, *unsystematic latent change* oder *systematic latent change* unterteilt. Ursprünglich beschreibt das LM-Modell letztgenanntes und ist in den *general LC framework* eingebettet. Ein *no latent change* Modell ist das *mixed* Markov (MM) Modell, welches eine Erweiterung des manifesten Markov-Modells in den *LC framework* darstellt und das Problem der Heterogenität behandelt. Die stationäre Variante des MM-Modells teilt die Stichprobe in eine Anzahl von latenten Klassen, wobei jede durch bedingte Antwortwahrscheinlichkeiten zum Zeitpunkt t_i und durch eine stationäre Transitionsmatrix charakterisiert ist. Zwei Fälle, wo die Matrix der Transitionswahrscheinlichkeiten der *stayer* gleich der Identitätsmatrix ist werden erwähnt. Einer dieser Spezialfälle des stationären MM-Modells ist das *mover-stayer*-Modell, wo die Transitionsmatrix der *mover* frei geschätzt wird. Der zweite Spezialfall ist das *black-and-white*-Modell, wo alle Transitionswahrscheinlichkeiten als auch die bedingten Antwortwahrscheinlichkeiten zum Zeitpunkt t_i gleichwahrscheinlich sind. Im stationären LM-Modell wird Stationarität sowohl für die bedingten Antwortwahrscheinlichkeiten, als auch für die Übergangsmatrizen angenommen. Das Modell kann als *error of measurement*-Modell gesehen werden, wobei die nicht bedingten Antwortwahrscheinlichkeiten als fehlerbehaftete Messungen angesehen werden, oder aber auch in Bezug auf die tatsächlichen latenten Klassen, wobei die nicht bedingten Antwortwahrscheinlichkeiten als wahre Antworten, welche mit einer gewissen Wahrscheinlichkeit auftauchen, angesehen werden, unabhängig davon, dass die Klasse durch ihre Antwortwahrscheinlichkeiten charakterisiert wird. Durch die nicht perfekte Messung, welche mit kategorialen Variablen einhergeht, wird vorrangig die messfehlerbehaftete Interpretation herangezogen.

Langeheine (1988) zeigt ein Drei-Klassen-LM-Modell wobei die Transitionsmatrix als Identitätsmatrix beschränkt wird. Dies kommt einem herkömmlichen stationären LC-Modell gleich, wobei die bedingten Antwortwahrscheinlichkeiten über die Zeit und Klassen gleich gehalten werden.

Langeheine (1988) erwähnt, dass die *Anderson-Goodman*-Entwicklungen in den manifesten Markov-Modellen durch die homogenen Prozesse, also eine Transitionsmatrix für alle Individuen, beschränkt sind. Die *mixed* Markov (MM) Modelle hingegen betrachten genau diese Heterogenität wobei eine unbekannt Anzahl an Segmenten von Individuen mit unterschiedlichen und möglicherweise nicht stationären Transitionswahrscheinlichkeiten eingebaut wird, also gleichzeitig Heterogenität, Stationarität und Ordnungseffekte löst. Ein Nachteil der MM-Modelle ist, dass ein Wechsel der Klassenzugehörigkeit über die Zeit nicht möglich ist. Die damals einzig bekannte Möglichkeit Veränderungen in der latenten Position zu messen ist das latente Markovketten-Modell, welches durch die bedingten Antwortwahrscheinlichkeiten Heterogenität auffängt, aber von Homogenität der latenten Transitionen ausgeht. Langeheine (1988) merkt an, dass vor der Unterstellung von Veränderungen das Modell zuerst der Nichtveränderungshypothese unterzogen werden muss. Weiters wird die Brauchbarkeit von Parameterbeschränkungen wie in den oben erwähnten stationären MM-Modellen, *mover-stayer* und *black-and-white*, auch für LM-Modelle vorgeschlagen. Auch Erweiterungen zur simultanen Analyse mehrerer Gruppen durch Einführung einer diskreten erklärenden Variablen in das Modell werden angedeutet.

Poulsen (1990) beschreibt das MM- und LM-Modell, welche Generalisierungen der LC-Modelle sind. Im MM-Modell wird die Homogenitätsannahme aufgeweicht und im LM Modell die Annahme der Zustandssicherheit. Die Kombination dieser Modelle ist das *latent mixed* Markov *chain*-Modell. Das MM-Modell ist eine finite Mischung (manifeste) Markovketten, welche individuelle Unterschiede in den Übergangswahrscheinlichkeiten erlaubt. Es weicht die Annahme lokaler Unabhängigkeit des LC-Modells auf. Das LM-Modell beschreibt Markovketten, welche auf der nicht beobachtbaren bzw. latenten Ebene basieren. Analysiert werden kategoriale Daten über die Benutzung latenter Variablen, wobei die Klassenzugehörigkeit der Individuen unbekannt ist, also *post hoc*, verglichen mit *a priori* bereits bekannten Klassenzugehörigkeiten. Die *zero-order*-Annahme der LC-Modelle wird durch Einführung von *first*- oder *higher-order*-Prozesse im *mixture* der Markovketten-Modelle aufgeweicht. Markov-Modelle mit Stadien, welche latente Veränderungsprozesse erlauben, werden eingeführt. Die latenten Modelle

ermöglichen die Analyse des Verhaltens und reduzieren die Störanfälligkeit der Daten durch Fokussierung auf stabile zugrundeliegende Strukturen wie Segmentzugehörigkeit und -profile. Markovketten haben einen schlechten Fit, da angenommen wird, dass alle Personen dem gleichen Veränderungsprozess folgen, also Gültigkeit der Transitionsmatrix für alle Individuen. Im MM-Modell wird Heterogenität in den individuellen Transitionswahrscheinlichkeiten erlaubt. Weiters wird angenommen, dass der Zustand des Prozesses ohne Fehler gemessen wird. Die Messung ist jedoch meist fehlerbehaftet und repräsentiert den wahren Zustand nicht genau. Eine Verbesserung wird durch LM-Modelle erreicht. So wie die *zero-order*-Modelle, entgehen die Modelle erster Ordnung nicht den Effekten der Populationsheterogenität, Ordnung und Nichtstationarität des Prozesses. Die *Anderson-Goodman*-Theorie ist auf homogene Prozesse beschränkt, also gleiche Transitionsmatrix für alle Individuen. Dies kommt der Annahme homogener *zero-order*-Prozesse in stochastischen Modellen gleich. Die *Brand Loyal*- und *Last Purchase Loyal*-Modelle beschreiben Modelle erster Ordnung mit zwei Zuständen und Heterogenität. Diese können jedoch nicht auf höhere Ordnungen erweitert werden und gehen von Stationarität aus.

Poulsen (1990) stellt eine Möglichkeit vor, die Population in Segmente zu unterteilen, welche unterschiedlichen nicht-stationären Prozessen erster Ordnung folgen. Dies kommt einer Generalisierung der LC-Modelle mit unterstellten *mixtures* von *zero-order* nicht-stationären *choice*-Modellen gleich. Das MM-Modell fängt Heterogenität, Ordnung und Nichtstationarität Individueller *choice*-Wahrscheinlichkeiten auf. Die gemeinsamen Wahrscheinlichkeiten bestehen aus Segmentgröße, anfänglichen Wahlwahrscheinlichkeiten und Transitionsmatrizen. Die Transitionsmatrizen beschreiben die bedingte Wahrscheinlichkeit für ein Individuum in einem bestimmten Segment zu einem bestimmten Zeitpunkt eine bestimmte Wahl, gegeben die vorangegangene Wahl, zu treffen. Das MM-Modell ist ein parametrisches multinomiales Modell.

4.3.7.28. Modellselektion

Als *goodness-of-fit* Bewertung zieht Poulsen (1990) die *LR-statistic* heran (*Likelihood-ratio* $L^2 = 2N \sum_i \sum_j \sum_k P_{ijk} \log(P_{ijk} / \hat{\pi}_{ijk})$).

Diese folgt einer asymptotischen χ^2 -Verteilung mit Freiheitsgraden. Die Freiheitsgrade berechnen sich aus der Anzahl der Zellen in der Tabelle minus eins minus der Anzahl an frei geschätzten Parametern. Einschränkungen durch Beschränkungen auf einen bestimmten Wert oder Gleichheitsbeschränkungen ohne einen bestimmten Wert können eingeführt werden. Wird von einem

ausgehenden Modell (M_1) eine Einschränkung in einem neuen Modell (M_2) eingeführt, so ist letztes genestet in ersterem. M_2 ist dann ein spezieller Fall von M_1 . Ist M_1 wahr, ist der Unterschied in der L^2 -goodness-of-fit statistic, $L^2(M_2 | M_1) = L^2(M_2) - L^2(M_1)$, eine bedingte Teststatistik für die Nullhypothese, dass die Beschränkungen von M_2 nicht signifikant sind. Unter dieser Hypothese wird $L^2(M_2 | M_1)$ asymptotisch einer χ^2 -Verteilung mit Freiheitsgraden $d.f.(M_2 | M_1) = d.f.(M_2) - d.f.(M_1)$ folgen. Beispiele sind das homogene Markov-Modell mit Segmentanzahl gleich eins oder das LC-Modell mit einer Wahrscheinlichkeit von null in einer der zwei Ausgangsklassen zu bleiben, weiters die Gültigkeit der *zero-order*-Annahme des LC-Modells für einige Segmente, aber nicht für andere, also Heterogenität in Bezug auf die Ordnung des Prozesses (Modellheterogenität), oder in Bezug auf die Parameter des Prozesses einer gegebenen Ordnung, weiters die Stationarität, wobei das *mover-stayer*- oder das *black-white-Modell mixtures* eines zwei-stationären Markov-Modells mit Werte- und Gleichheitsbeschränkungen der Transitionsmatrix sind. Auch *mixtures* höherer Ordnung können als werte- und gleichheitsbeschränkte Modelle erster Ordnung angesehen werden. Letztendlich wird noch die simultane MM-Analyse für Gruppenanalysen von Poulsen (1990) erwähnt.

4.3.7.29. Übergänge zwischen unterschiedlichen Populationen

MM-Modelle, als auch LC-Modelle, gehen von der Annahme aus, dass die Segmentzugehörigkeit jedes Individuums fix ist. Der beobachtete Wechsel reflektiert die probabilistische Natur des individuellen Auswahlprozesses und nicht die darunterliegende Präferenzstruktur. Alle Individuen eines Segmentes folgen dem gleichen *zero-order*-Prozess im LC-Modell, oder einem Prozess erster Ordnung im MM-Modell. Es kann keine richtigen Veränderungen der Klassenzugehörigkeit geben und somit kann eine Klasse nicht wachsen oder schrumpfen. Poulsen (1990) kombiniert deshalb die stochastische Beschreibung der individuellen Auswahl mit Veränderungen im individuellen Prozess. So wird ein finites Set an Zuständen im Markov-Modell definiert. Die Zustände sind stationäre *zero-order choice*-Modelle und Wechsel zwischen diesen Zuständen basieren auf einer stationären Übergangsmatrix. Da weder Zustände noch Übergangsmatrix des Markov-Modells direkt beobachtbar sind, aber aus den manifesten *choices* abgeleitet werden müssen, wird das Modell LM-Modell genannt. Dieses berücksichtigt die Unsicherheit in der Messung jedes individuellen Zustandes in der Markov-Modellierung. Jeder Zustand wird durch einen stationären Wahrscheinlichkeitsvektor über die *choice*-Sets beschrieben, welche die strukturelle Relation zwischen den manifesten *choices* und den latenten Zuständen auffängt. Das LM-Modell besteht aus den Parametern eines

multinomialen Modells, also die anfänglichen Zustands-, die Auswahl- und die Übergangswahrscheinlichkeiten. Somit besteht das Modell erstens aus der probabilistischen Relation zwischen den manifesten beobachteten Antworten und den latenten nicht beobachtbaren wahren Zuständen und zweitens aus dem Veränderungsprozess, welcher auf den latenten Zuständen basiert und durch die Markov-Übergangsmatrix beschrieben wird. Werden Markovketten-Modelle als *brand choice*-Modelle eingeführt, können diese als Generalisierungen der *zero-order*-Modelle mit Einfluss vorangegangener *choices* gesehen werden. Diese Betrachtung übersieht die strukturelle Interpretation der *zero-order*-Modelle, wo die beobachteten *choices* als nicht perfekte Indikatoren oder Manifestationen der zugrundeliegenden nicht beobachtbaren Wahrscheinlichkeiten gesehen werden. Sind die *choices* reine Realisierungen eines stochastischen Prozesses, kann argumentiert werden, dass die Markov-Modelle, welche Veränderungsprozesse beschreiben, eher auf stabileren stochastischen Prozessen basieren sollten, als auf zufälligen Realisierungen. Dies führt eher zur Definition des Zustandsraumes in Sinne von latenten *choice*-Prozessen als zu manifesten *choices*.

Beschränkte LM-Modelle können für bedingte Tests verwendet werden. Das Nichtvorhandensein der Zustandsunsicherheit, also dass die *choices* perfekte Indikatoren der zugrundeliegenden Zustände sind, kann überprüft werden. Dies kommt wiederum dem traditionellen homogenen Markov-Modell gleich. Die Anzahl der latenten Zustände entspricht dabei der Anzahl der *choice* Alternativen. Auch das Nichtvorhandensein einer latenten Veränderung kann überprüft werden. Dabei wird der beobachtete Wechsel zwischen Alternativen durch die Zustandsunsicherheit erklärt, wobei keine wirklichen Veränderungen in der latenten Position auftauchen, was zum LC-Modell mit *parallel profiles* führt.

4.3.7.30. *Mixed Markov latent class model*

Van de Pol und Langeheine (1990) beschreiben *mixed Markov latent class* (MMLC) Modelle. Nur selten kann davon ausgegangen werden, dass jeder Respondent einer Panelstudie der gleichen Veränderung einer diskreten Charakteristik folgt. *Discrete time* MMLC-Modelle beschreiben nicht den Veränderungsprozess zwischen Beobachtungen, wie dies kontinuierliche Zeitmodelle tun. MMLC-Modelle beschreiben Kreuztabellen aufeinanderfolgender Messungen anhand mehrerer Markovketten. So werden diese Modelle für mehrere Subgruppen zur simultanen Analyse erweitert. Die Population wird aufgrund einer Charakteristik aufgesplittet, welche relevant für die Übergangswahrscheinlichkeiten oder andere Parameter ist. Dadurch steigt der Einblick in die Heterogenität des Prozesses. Es werden einzelne Markovketten,

das *mover-stayer*-Modell, das *black-and-white*-Modell, das LM-Modell, Markov-Modelle mit Übergängen zwischen Ketten, multivariate MMLC-Modelle, Markovketten höherer Ordnung und MMLC mit simultaner Analyse mehrerer Subpopulationen vorgestellt, wobei Parameter fixiert oder gleichgesetzt werden. Gleichheitsrestriktionen von Parametern über Subgruppen werden über LRT getestet.

Es geht also um Realisierungen gleicher polytomer Variablen aufeinanderfolgender Messungen. Ein Extremfall der MMLC-Modelle ist ein Modell, welches die gesamte beobachtete Veränderung den Messfehlern beimisst. Ein derartiges *no-latent-change*-Modell ist das LC-Modell. Jeder Respondent in einer Klasse hat die gleichen Antwortwahrscheinlichkeiten einer Antwort zu einem Zeitpunkt. LC-Modelle gehen davon aus, dass jede manifeste Variable eine nicht perfekte Messung des selbigen zugrundeliegenden statischen Faktors ist. Im *single* Markovketten-Modell ist lediglich die jüngste Beobachtung zur Vorhersage des derzeitigen Zustandes relevant. Es geht von einer homogenen Population aus. Jede Person in einem Zustand zu einem Zeitpunkt hat die gleiche Übergangswahrscheinlichkeit einen gewissen Zustand zum darauffolgenden Zeitpunkt einzunehmen, unabhängig vom vorangehenden Zeitpunkt. Die Tatsache den gleichen Zustand zu zwei Beobachtungen anzunehmen ist keine Garantie dafür, dass es keine Veränderung in der dazwischen liegenden Periode gegeben hat. Kontinuierliche *time models* beziehen auch diese Veränderung mit ein, diskrete *time models* wie MMLC-Modelle nicht. Populationsheterogenität wird durch die Annahme mehrerer Markovketten modelliert und Nichtreliabilität durch die Definition latenter Markovketten.

4.3.7.31. *Mover vs. stayer*

Drei spezielle Fälle von *single* Markovketten werden erwähnt. Bei Unabhängigkeit, *independence*, ist der derzeitige Zustand unabhängig vom vorhergehenden Zustand. Die Übergangswahrscheinlichkeitsmatrix geht von gleichen Zeilen und gleicher marginaler Verteilung aus. Diese können sich zwischen den Zeitpunkten unterscheiden. Bei Stationarität (*stationarity*) sind die Übergangswahrscheinlichkeiten über die Zeit homogen. Ist eine *independence*-Kette stationär, so ist die Wahrscheinlichkeit jedes Zustandes, also die marginale Verteilung für jeden Zeitpunkt gleich. Die Parameter sind in jeder Übergangsperiode gleich. Wird eine zeitlich homogene marginale Verteilung angenommen, sind die Übergangswahrscheinlichkeiten gleich der marginalen Verteilung zu jedem Zeitpunkt. Bei keiner Veränderung (*no change*) existieren die gleichen Kategorien für jede Beobachtung ohne latente Veränderung. Die Wahr-

scheinlichkeit, den selbigen Zustand wie zuvor anzunehmen ist eins. Diese *no-change-* oder *stayer-*Übergangsmatrix ist gleich der Identitätsmatrix.

Die beschränkte Zugehörigkeit zu einer Kette erlaubt keine Übergänge zwischen Ketten. Das *black-and-white-* und *mover-stayer-*Modell sind MM-Modelle mit zwei Ketten und zusätzlichen Restriktionen. Im *mover-stayer-*Modell gibt es eine nicht beschränkte Kette, die *mover-*Kette, und eine beschränkte Kette, die *stayer* Kette. Das *black-and-white-*Modell ist ein *mover-stayer-*Modell mit einer *mover-*Kette, welche aus unabhängigen Antworten, *independent responses*, oder sogar zufälligen Antworten (*random response*) besteht. Weitere beschränkte MM-Modelle mit zwei Ketten liegen in der Kombination von *movers* und Personen deren Antworten unabhängig von der vorangegangenen sind, *mover-independence-*Modell, oder eine *mixture* von *movers* und Personen welche zufällig antworten, *mover-random-response-*Modell.

Das LM-Modell ist eine Kombination aus einem nicht beschränkten LC-Modell und einer einzelnen Markovkette. Die gesamte Population verändert sich entsprechend einer Markovkette. Messfehler bewirken Abweichungen der beobachteten Kreuztabellen, von dem was beobachtet werden würde, wenn die Reliabilität perfekt wäre. Die Antwortwahrscheinlichkeiten der ersten und der letzten Beobachtung können nicht aus den beobachteten Daten ohne weitere Restriktionen für das LM-Modell geschätzt werden. Oft erscheint es sinnvoll, für drei Zeitpunkte, diese gleich der nächstgelegenen Beobachtungen zu setzen. Für vier oder mehr Beobachtungen müssen keine zeitlich homogenen Antwortwahrscheinlichkeiten angenommen werden. Werden die Übergangswahrscheinlichkeiten gleichgesetzt, reduziert sich das LM-Modell zum latenten *stayer-*Modell ohne Übergangswahrscheinlichkeiten wie im LC-Modell. Eine *mixture* mehrerer latenter Markovketten, das *latent mixed* Markov (LMM) Modell, wird unter Annahme der Nichtreliabilität gemessen. Das MM-Modell ist ein Spezialfall des LMM, wenn die Anzahl der latenten Klassen gleich der Anzahl an manifesten Kategorien ist und die Reliabilität perfekt ist. Alle bis dahin von Van de Pol und Langeheine (1990) bereits erwähnten Modelle sind Spezialfälle des LMM. Um klarzustellen, dass diese Modelle LC-Modelle sind, beziehen sie sich auf das MMLC-Modell, wobei LMM als Name für das LM-Modell steht, welches beschränkt als MM auf einem latenten Level ist. Restriktionen um das LMM-Modell zu identifizieren, werden über Beschränkung zu einem *mover-stayer-*Modell mit perfekter Reliabilität für die *stayers* eingebracht.

Der Fit genesteter Modelle wird über LRT verglichen. MM-Modelle, mit Ausnahme des *Zwei-Ketten-mover-stayer*-Modells und Submodelle, müssen mit mehreren Sets an Startwerten gerechnet werden, da diese Modelle lokale Maxima aufweisen können. LRT wird verwendet, um die Stationarität einer Markovkette zu zeigen, indem der Fit einer nicht stationären Markovkette mit dem einer stationären verglichen wird.

4.3.7.32. Latent Markov model - Spezifikation

Das MM-Modell besagt, dass die Personen aus allen möglichen Ketten durchgehend der gleichen latenten Kette über die Zeit angehören. Dies ist unrealistisch. Deshalb sollte ein Übergang zwischen Ketten erlaubt werden. Das Modell wird zu einem Modell mit Übergängen zwischen allen latenten Klassen erweitert. Die Null-Restriktionen, in Diagramm 141 auffindbar, werden zumindest teilweise entfernt.

Structural Zeroes in Response Probabilities, ρ_{iia} , and Transition Probabilities, τ_{iia} , of the Mixed Markov Model, Formulated as a Latent Markov Model

	Response Probabilities		Transition Probabilities			
			b=1	b=2	b=3	b=4
	i=1	i=2	j=1	j=2	j=1	j=2
a=1, chain 1, category 1 (i=1)	1	0	$\tau_{...}$	$\tau_{...}$	0	0
a=2, chain 1, category 2 (i=2)	0	1	$\tau_{...}$	$\tau_{...}$	0	0
a=3, chain 2, category 1 (i=1)	1	0	0	0	$\tau_{...}$	$\tau_{...}$
a=4, chain 2, category 2 (i=2)	0	1	0	0	$\tau_{...}$	$\tau_{...}$

Diagramm 141: Mixed Markovmodell formuliert als latentes Markovmodell (Van de Pol und Langeheine, 1990)

Markovketten höherer Ordnung weichen die Markovianische Annahme eines Prozesses ohne Gedächtnis auf. So hängen die Übergangswahrscheinlichkeiten in der dritten Beobachtung vom Zustand bei der ersten und vom Zustand zum zweiten Beobachtungszeitpunkt ab.

Auch Modelle für mehrere Subpopulationen sind denkbar. Die Samplegröße für die Subpopulationen ist dann kleiner. Komplexere Modelle werden über Parametereinschränkungen, welche sich als gleich über die Subpopulationen dargestellt haben, gefittet. So wird das LMM-Modell um den Parameter des Subpopulationsanteiles erweitert und alle anderen Parameter sind bedingt durch

die Zugehörigkeit zur Subpopulation. Die Parameter können zwischen jeder Subpopulation unterschiedlich sein.

Ein Modell kann theoriebasiert ausgewählt werden, wobei hier jedoch das Risiko vorliegt, dass ein gut passendes Modell abgelehnt wird, da es nicht den theoretischen Annahmen entspricht. Umgekehrt kann ein rein empirischer Ansatz, welcher den Fitvergleich aller Kandidaten über LRT vornimmt, nicht blind ohne Geltung der Bedeutung des Modells verwendet werden.

Van de Pol und Langeheine (1990) stellen einen Weg zur Auswahl von Gleichheitsbeschränkungen über Subpopulationen eines gegebenen Modells vor. Zuerst werden zwei extreme Spezifizierungen geschätzt, wobei in einer alle Parameter über die Subpopulationen gleich sind und in der anderen alle Parameter zwischen den Subpopulationen unterschiedlich sind. Dies erlaubt einen Benchmark, welche Verbesserung des Fits durch Aufweichung der Gleichheitsbeschränkungen möglich ist. Die Lösung mit dem besten Fit wird durch die Aufweichung der Gleichheitsbeschränkungen auf einem systematischen hierarchischen Weg gesucht, zuerst der Populationsanteil für die Markovketten, dann der Anteil der unterschiedlichen Klassen, dann die Antwortwahrscheinlichkeiten und schließlich die Zustände. Diese Restriktionen werden eine nach der anderen entfernt, z.B. alle Parameter für die Population in den Markovketten, und im Anschluss jeweils die Fitverbesserung über LRT evaluiert. Wird keine signifikante Verbesserung erzielt, wird die Prozedur verfeinert, indem keine Sets, sondern individuelle Parameter betrachtet werden. Wird keine signifikante Verbesserung mehr erzielt, stoppt der Prozess. Wird der Fit signifikant verbessert, wird die meist restriktive Gleichsetzungsbeschränkung aufgeweicht, und bedingt durch diese Annahme, wird die Prozedur wiederholt und die nächst bessere Verbesserung des Fits wird gesucht.

Continuous-time Parameter hängen nicht von der Länge der Perioden zwischen den Panelwellen ab, *discrete-time* Parameter hingegen schon. So wurden auch *continuous-time* Versionen von LM-Modellen entwickelt. Das *discrete-time mixed* (Markov) Modell wird auch nichtparametrisches Modell genannt und MMLC Modelle haben den Vorteil, dass keine restriktiven Verteilungsannahmen gemacht werden müssen.

Van de Pol und Mannan (2002) beschreiben die Analyse kategorialer Daten in diskreter Zeit mit Hilfe von LM Modellen und geben Hinweise zum Schätzprogramm PANMARK. LCA betrachtet kategoriale Daten über Häufigkeitstabellen. Sie wird für die Datenreduktion von *cross-sectional* Daten

oder für wiederholte Messungen unter Annahme von Markovketten verwendet. Mit mehreren Indikatoren ermöglicht ein LC-Modell die Schätzung der Fehlklassifikationswahrscheinlichkeiten und Häufigkeitstabellen der latenten (*hidden*) Variablen hinter den gemessenen Indikatoren. Durch Miteinbeziehung der Messfehler erhält man ein genaueres Bild über die Beziehung zwischen den latenten Variablen und den beobachteten Häufigkeitstabellen. Messfehler und wahre Veränderung werden durch die Annahme von Markov-Prozessen separiert. Populationsheterogenität kann rein aufgrund von Hintergrundcharakteristiken meist nicht vollständig erklärt werden. Dieser nicht erklärte Teil wird latente Heterogenität genannt, wie beispielsweise das *mover-stayer*-Modell oder anspruchsvollere *mixtures* von Markovketten. Weiters werden homogene Cluster von Objekten lokalisiert. Durch Tabellenreduktion entsteht Datenreduktion. Tabellen, welche eine große Anzahl an Teilstichproben beschreiben, werden auf eine kleine Anzahl latenter Tabellen reduziert. Die Ladungen der Teilstichproben auf die latenten Tabellen werden generiert. So beschreiben Kreuztabellen eine große Anzahl an Kohorten und durch die simultane Analyse dieser Kohorten werden beispielsweise zwei latente Tabellen geschätzt. Dies ist für Kohorten, welche aufgrund von Panelmortalität entstehen, brauchbar.

Van de Pol und Mannan (2002) listen ausführlich die Notation der MMLC-Modelle auf. Die beobachteten Variablen sind Indikatoren latenter Variablen, welche quasi indirekt gemessen werden und auch *hidden* Variable genannt werden. Zwecks Modellidentifikation ist die Beschreibung der Entwicklung der latenten Variablen durch eine Markovkette nützlich. Durch *mixtures* von Markovketten kann mehr als ein Entwicklungstyp modelliert werden. Um diese Modelle zu identifizieren, sind mehrere Indikatoren für die gleiche latente Variable nützlich. Kategorische exogene Variable können zur Beschreibung der Subpopulationen hilfreich sein. MMLC ist auf die Aufdeckung unterschiedlicher Subpopulationstypen fokussiert. Jedes Subjekt gehört einer Subpopulation an. Die Zugehörigkeit zu dieser Population wird als konstant über die Zeit angenommen. So hat jede Subpopulation einen Anteil. Alle anderen Parameter sind bedingt durch die Subpopulation und somit können alle oder auch nur einige der Parameter unterschiedlich in den Subpopulationen sein. Jedes Mitglied einer Subpopulation gehört einer latenten oder manifesten Kette von Personen mit derselben Dynamik an. Der Anteil der Subpopulation gehört zur Kette. Ein Mitglied einer Subpopulation und Kette gehört zu einer Klasse. Somit gibt es einen Anteil der Klasse für eine Subpopulation bzw. Kette. Für die Wahrscheinlichkeit einer Indikatorantwort, gegeben die Subpopulation bzw. Kette und Klasse, wird angenommen, dass diese für alle Subjekte in der Subpopulation der

Kette und Klasse gleich ist. Ist die Variable des Modells nicht latent aber manifest, so entspricht die latente Klasse den manifesten Kategorien. Die Antwortwahrscheinlichkeiten werden damit im manifesten Modell überflüssig. Hat das Modell keine latente Variable zum Zeitpunkt eins und zwei, wird angenommen, dass für die Subpopulation jedes Mitglied der Kette die gleichen Übergangswahrscheinlichkeiten von einer Klasse zum Zeitpunkt eins zu einer Klasse zum Zeitpunkt zwei aufweist. So wie für den Indikator zum Zeitpunkt eins, als auch für den Indikator zum Zeitpunkt zwei, wird die Wahrscheinlichkeit einer gewissen Antwort gegeben Klasse, Kette und Subpopulation, angenommen, dass diese für alle Subjekte einer Subpopulation einer Kette und Klasse gleich ist. Die Erweiterung für mehrere Indikatoren erfolgt sinngemäß. Ein *partially* LMM-Modell ist beispielsweise das *mover-stayer* Modell mit Antwortfehlern nur für die *movers*. Keine Antwortfehler für die *stayers* anzunehmen ist plausibel und praktisch.

4.3.7.33. Modelleigenheiten und Kovariate

Die erste Beobachtung im Panel weist oft geringere Reliabilität auf, verglichen mit den folgenden. Soll dieses Phänomen unterschiedlicher Antwortwahrscheinlichkeiten zum ersten Zeitpunkt untersucht werden, können stationäre Übergangswahrscheinlichkeiten angenommen werden. LR ist ein Modellfitkriterium welches durch χ^2 -Verteilung evaluiert wird, solange die Stichprobengröße in jeder Zelle der Kontingenztabelle groß genug ist. Ein weiteres Kriterium dieser theoretischen Verteilung sind die Freiheitsgrade. Manchmal werden Parameter im Zuge der Schätzung auf die Grenzen des Parameterraumes, null und eins, fixiert. Der Effekt dieser Grenzwerte auf die Freiheitsgrade ist unsicher. Gleichheitsbeschränkungen verbieten die Annahme einer einfachen theoretischen χ^2 -Verteilung, da durch Gleichheiten eine kompliziertere Verteilung entsteht. Wird eine dieser Bedingungen nicht erfüllt, kann die Verteilung der *Fit-statistic* mit dem *repeated parametric bootstrap sample* generiert werden. LC-Modelle weisen mehr als ein Optimum der *likelihood*-Fläche auf. Die optimale wird bevorzugt. Die Optimierungsroutine endet jedoch oft in einem lokalen Optimum. Deshalb wird empfohlen, den Algorithmus mit mehreren Iterationen von mehreren Punkten der *likelihood*-Fläche zu starten und mit der vielversprechendsten fortzufahren. Je mehr Parameter involviert werden, desto mehr Startpunkte werden benötigt. Die Beziehung ist nicht linear, sondern erscheint exponentiell. Wird die Anzahl unabhängiger Parametern verdoppelt, sollte die quadrierte Anzahl an Startpunkten verwendet werden, was jedoch nicht praktikabel ist.

Auch exogene Variable, also Kovariate, können latente Parameter erklären. Es gibt zwei Schulen zur Einbringung erklärender exogener Variablen in das Messmodell. Eine bevorzugt die simultane Schätzung aller Parameter. Andererseits sollten die erklärenden Variablen keinen Einfluss auf das Messmodell ausüben. Eine Variable, welche stark mit einem Indikator im Messmodell zusammenhängt, verstärkt die Reliabilität der Schätzung dieses Indikators und die Reliabilität der anderen Indikatoren erscheint folglich schwächer als sie tatsächlich ist. Die Vorteile der simultanen Schätzung liegen im Nichtvorliegen der Gültigkeit des Messmodells für die gesamte Population. Van de Pol und Mannan (2002) erläutern ein Beispiel. Ist die Reliabilität der männlichen Raucher geringer als die der weiblichen und wird dieser Unterschied im Messmodell ignoriert, wird der Geschlechtereinfluss mit einem Fehler geschätzt. Dieser Typ an Heterogenität kann durch die simultane Schätzung der Messmodelle in den Subgruppen der Stichprobe getestet werden.

4.3.7.34. Modellhierarchie

Langeheine und Van de Pol (2002) listen zwei Unterschiede in der Analyse longitudinaler kategorialer Daten auf. Für wenige Subjekte, welche über viele Beobachtungszeitpunkte beobachtet wurden, bieten sich Zeitreihenanalysen an. Für viele Subjekte, welche über wenige Beobachtungszeitpunkte gemessen wurden, bieten sich Markovketten-Modelle an. Diese beziehen sich auf diskrete Variable, also kategoriale Items bzw. Indikatoren mit Antwortkategorien. Sowohl die manifesten, als auch die latenten Variablen sind kategorial, also diskret. Weiters besitzen Markovketten-Modelle die Annahme über diskrete Zeit zu untersuchen, also Aussagen über den Übergang von einem Zeitpunkt zum nächsten. Im Unterschied zu kontinuierlichen Zeitmodellen beschreiben sie nicht den Veränderungsprozess zwischen Beobachtungen. Das übergeordnete Modell ist das LMM-Modell für mehrere Gruppen. Alle darunterliegenden Modelle lassen sich, wie in Diagramm 142 dargestellt, davon ableiten.

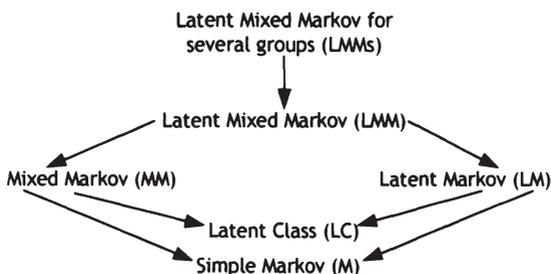


Diagramm 142: Markovkettenmodellhierarchie (Langeheine und Van de Pol, 2002)

Im *simple* Markov Modell hängt die Wahrscheinlichkeit, sich in einem gewissen Zustand zu befinden nur vom vorangehenden Zeitpunkt ab. Es geht von Homogenität der Population aus. Veränderungen sind für alle Subjekte gültig. Wird dies zum MM-Modell erweitert, kann jede Kette ihrer eigenen Veränderung folgen. Weiters nimmt das *simple* Markov-Modell an, dass die Beobachtungen frei von Messfehlern sind. Wird dieses zum LM-Modell erweitert, korrigiert dies den Messfehler. Das LC-Modell ist ein Spezialfall sowohl des *mixed* als auch des latenten Markov-Modells. Dafür wird im LM-Modell keine Veränderung angenommen, also die Transitionsmatrizen der Einheitsmatrix gleichgesetzt. Im MM-Modell wird angenommen, dass die Übergangswahrscheinlichkeiten zu einem Zeitpunkt nicht vom vorangehenden Zeitpunkt abhängen. Anstatt also Abhängigkeit zwischen aufeinanderfolgenden Zeitpunkten anzunehmen, wird lokale Unabhängigkeit wie im klassischen LC-Modell angenommen. Das LM Modell hat nur eine Kette, die Erweiterung zum LMM-Modell erlaubt mehrere Ketten und nicht beobachtete Populationsheterogenität auf latentem Level. Sind die Messungen perfekt, reduziert sich dies zum MM-Modell, also mehrere Ketten mit perfekten Messungen. Die Erweiterung zum LMM für mehrere Gruppen, ausgehend vom LMM-Modell, inkludiert die Situation, Daten von mehreren Gruppen zu haben, welche durch zusätzliche externe diskrete Variablen für die simultane Analyse aller Gruppen eingebracht werden.

Langeheine und Van de Pol (2002) beschreiben den Aufbau vom simpelsten Modell weg. Im *simple* Markovketten-Modell wird ein Prozess erster Ordnung angenommen. Da außer dem vorangehenden Zeitpunkt keine weiter zurückliegenden miteinbezogen werden, wird dies auch *process without memory* genannt, in Diagramm 143 dargestellt. δ^t beschreibt die marginalen Verteilungen zu den jeweiligen Zeitpunkten, T die homogenen Transitions-wahrscheinlichkeiten. Die Brauchbarkeit dieses und auch der folgenden Modelle, kann zusätzlich zur Devianz und BIC, anhand des *Dissimilarity*-Index, welcher die Anzahl der Missspezifizierungen, also Falschzuordnungen bewertet, beurteilt werden. Werden die marginalen Verteilungen einzeln zu den Zeitpunkten berechnet, kann bei einer Tendenz zu keiner Veränderung auf ein Equilibrium des Systems in Richtung stabilen Zustand identifiziert werden.

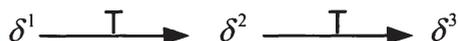


Diagramm 143: Classic Markov chain Modells (Langeheine und Van de Pol, 2002)

Im MM-Modell existieren Subgruppen mit spezifischen Prozessen über die Zeit. Dabei werden *intragroup*-Homogenität und *intergroup*-Unterschiede maximiert. Zu Beginn existierten Vorschläge, die Stichprobe anhand einer oder mehrerer exogener Variablen aufzusplitten und ein *simple* Markov-Modell für jede der Gruppen anzupassen. Der Fokus liegt jedoch auf der nicht beobachteten Heterogenität. Eines der ersten Modelle ist das *mover-stayer*-Modell, wobei die *mover* einer gewöhnlichen Markovkette folgen, die Übergangswahrscheinlichkeitsmatrizen der *stayer* jedoch gleich der Identitätsmatrix sind. Das Ziel liegt in der Subgruppenaufteilung der beobachteten Häufigkeitsverteilungen, welche die Schätzung der Gruppenanteile und Parameter der Subgruppen, unter der Annahme einer finiten Anzahl an Markovketten erlaubt. Somit existieren Parameter für die Kettenanteile, wobei alle weiteren Parameter bedingt durch die Kettenzugehörigkeit sind, im Falle ohne Messfehler (manifeste Ketten). Anstatt der Annahme lokaler Abhängigkeit zwischen aufeinanderfolgenden Zeitpunkten innerhalb einer Kette, einem Markovprozess erster Ordnung, besteht lokale Unabhängigkeit, oder ein *zero-order*-Prozess innerhalb der Klassen im Modell. Konzepte wie *mover*, *stayer*, *independence* und *random* können beliebig kombiniert werden.

Sind die Zeilen der Transitionsmatrix annähernd gleich, wird davon ausgegangen, dass der Prozess einen stabilen Zustand erreicht hat. Die Nettoveränderung wird aus der Differenz zwischen anfänglicher Verteilung, also dem Anteil der Klasse eins, und jenem Anteil an Personen, welche zum letzten Zeitpunkt in der gleichen Klasse verbleiben, abgelesen.

4.3.7.35. Interpretationshinweise - Kinodatensatz

Langeheine und Van de Pol (2002) vergleichen ein *mover-stayer*- mit einem MM-Modell mit zwei Klassen. Sie weisen darauf hin, dass ein kleiner Übergang einer großen Klasse, einen großen Übergang einer kleinen Klasse aufwiegen kann. Deshalb dürfen nicht rein die Übergänge betrachtet werden, sondern das Produkt aus dem Anteil der jeweiligen Klasse, multipliziert mit dem Anteil des Überganges. Diese Situation wird in Tabelle 96 anhand des Drei-Klassenmodell der zuvor vorgestellten LTA mit stationären Übergangswahrscheinlichkeiten dargestellt. Die Übergangswahrscheinlichkeiten zwischen den latenten Zuständen zeigen, dass die Wahrscheinlichkeit in der gleichen Klasse zu verbleiben, am höchsten für die Klassen eins und zwei ist, 93,9 bzw. 96,2%. In Klasse eins und zwei existieren lediglich geringe Übergangswahrscheinlichkeiten zu anderen Klassen. Klasse drei zeigt die höchsten Übergangswahrscheinlichkeiten. Die meisten Kinogeher der Klasse drei hüpfen in Klasse zwei.

		Zeitpunkt <i>t</i>		
		Class 1	Class 2	Class 3
Zeitpunkt <i>t-1</i>	Class 1	0.939	0.005	0.057
	Class 2	0.012	0.962	0.025
	Class 3	0.239	0.632	0.129

Tabelle 96: Stationäre Übergangswahrscheinlichkeitsmatrix

Jedoch müssen die anfänglichen Klassenwahrscheinlichkeiten berücksichtigt werden. In Tabelle 97 werden die Übergänge der Klassen zwischen Zeitpunkt eins und zwei in absoluten Zahlen exemplarisch interpretierbar gemacht. Die Übergangswahrscheinlichkeiten müssen dafür mit den anfänglichen Klassenwahrscheinlichkeiten multipliziert werden.

Transitions	Calculation	Size
Class one to class one	13.997% x 93.9%	13.143180%
Class one to class two	13.997% x 0.5%	0.069985%
Class one to class three	13.997% x 5.7%	0.797829%
Class two to class one	82.91% x 1.2%	0.994920%
Class two to class two	82.91% x 96.2%	79.759420%
Class two to class three	82.91% x 2.5%	2.072750%
Class three to class one	3.092% x 23.9%	0.738988%
Class three to class two	3.092% x 63.2%	1.954144%
Class three to class three	3.092% x 12.9%	0.398868%

Tabelle 97: Absolute Übergangswahrscheinlichkeiten

Die meisten Respondenten, 79,76% verbleiben in der zweiten Klasse zwischen Zeitpunkt eins und zwei, 13,14% in der ersten Klasse. Respondenten, welche von Klasse zwei in Klasse drei hüpfen, 2,07% und jene, die von Klasse drei nach Klasse zwei hüpfen, 1,95%, stellen die höchsten Übergänge zwischen den weniger zufriedenen Klassen dar. Würde man lediglich die Übergangswahrscheinlichkeiten zwischen diesen Klassen interpretieren, 63,2% und 2,5%, so würde der höchste Übergang von Klasse drei nach zwei passieren. Da die Klasse drei aber sehr klein ist, würden die Respondenten zu stark gewichtet werden. Die latenten Klassenzahlen des geschätzten Modells in Tabelle 98 verdeutlichen dieses Argument. Klasse drei ist die kleinste Klasse am unteren Zufriedenheitslevel. Klasse eins befindet sich in der Mitte des latenten Zufriedenheitsraumes zwischen den Klassen drei und eins. Die weniger zufriedenen Klassen eins und drei gewinnen an Respondenten über die Zeit, in Summe 77 Respondenten mit 8,82% aller 1.737 Respondenten. Klasse zwei

befindet sich am oberen Ende des Zufriedenheitskontinuums. Verliert sie an Respondenten, so tendieren die Respondenten, welche auf diesem latenten Zustand starten, in Richtung niedriger latenter Zufriedenheitszustände. Die Übergangswahrscheinlichkeiten in Klasse zwei zeigen zwar sehr niedrige Werte, 0,5 bzw. 5,7%, aufgrund der großen Anzahl an Personen in dieser Klasse macht sich jedoch ein sinkender Zufriedenheitstrend bemerkbar.

	Time point 1	Time point 2	Time point 3	Time point 4	Time point 5	Time point 6
Class one	243	259	274	288	301	314
Class two	1,440	1,421	1,405	1,390	1,376	1,363
Class three	54	57	58	59	59	60

Tabelle 98: Endgültige Klassenzahlen

4.3.7.36. Verschachtelung und Modellidentifikation

Ein *mover-mover-stayer*-Modell mit nicht-stationären Übergangswahrscheinlichkeiten ist genestet im allgemeineren *mover-mover-mover*-Modell und kann über die *likelihood*-Differenz verglichen werden. Werden Übergangswahrscheinlichkeiten als stationär angenommen, wird BIC *mover-mover-stayer*- und *mover-mover-mover*-Modelle aufgrund der Sparsamkeit der geschätzten Parameter bevorzugen. Probleme der MM-Modelle liegen in der Identifizierung, weshalb unterschiedliche Startwerte getestet werden sollen. Weiters gehen MM-Modelle davon aus, dass die Zugehörigkeit zu einer Kette über die Zeit konstant ist, was einer Vereinfachungsannahme des zugrundeliegenden stochastischen Prozesses gleichkommt. Das MM-Modell erlaubt also nicht die Evaluierung der wahren Veränderung, also den Übergang von einer zu einer anderen Kette über die Zeit.

Sowohl die Messfehler als auch die latente Veränderung sind Elemente der LM-Modelle. Im LM-Modell startet der Prozess mit einer latenten Verteilung, mit latenten Zuständen bzw. Klassen, zu welchen die manifesten Antworten der Subjekte probabilistisch verbunden werden, die Antwortwahrscheinlichkeiten. Der obere Teil aus Diagramm 144 zeigt das LM-Modell, der untere das LC-Modell. u' sind die manifesten marginalen Verteilungen zum Zeitpunkt t , welche mit den latenten Verteilungen δ' durch die Antwortwahrscheinlichkeiten R' verbunden werden. Die Abweichung von R zur Einheitsmatrix beschreibt die Messfehler. Also verbindet das Messmodell R die latenten Variablen mit den manifesten. Das Strukturmodell ist durch δ und T gegeben.

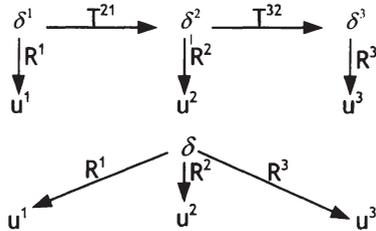


Diagramm 144: Latentes Markov Modell vs. latentes Klassenmodell (Langeheine und Van de Pol, 2002)

Solange keine weiteren Restriktionen eingeführt werden, ist das simple Markovketten-Modell nicht identifizierbar. Beispielsweise können Antwort- und Übergangswahrscheinlichkeiten zeithomogen gesetzt werden. Die Übergangswahrscheinlichkeitseinschränkung ist für die Identifizierung nicht unbedingt notwendig, die Antwortwahrscheinlichkeitsbeschränkung im Falle von drei Zeitpunkten jedoch schon. Für mehrere Zeitpunkte müssen lediglich zeithomogene Antwortwahrscheinlichkeiten für den ersten und den letzten Zeitpunkt angenommen werden. Ein stabiles Messmodell ist zwecks Interpretation auch leicht rechtfertigbar. Weist das LM-Modell einen guten Fit auf, beweist dies keine latente Veränderung, solange das Modell nicht gegen das Modell ohne latenter Veränderung, also alle Transitionsmatrizen gleich der Einheitsmatrix, getestet wurde, was dem LC-Modell gleichkommt. Im LM-Modell werden Veränderungen nicht auf manifestem, sondern latentem Level beschrieben, *true change* genannt.

4.3.7.37. Modellvergleich - Kinodatensatz

Bevor Veränderungen den Kinodatensatz betreffend angenommen werden und die eben theoretisch erläuterten Modelle empirisch getestet werden, müssen die zu Beginn des Kapitels gerechneten Modelle erst mit einem Modell ohne Veränderungsannahme verglichen werden. Diese entsprechen Modellen, in welchen die Übergangswahrscheinlichkeiten zwischen unterschiedlichen latenten Klassen gleich null gesetzt werden. Also die Übergangswahrscheinlichkeitsmatrix einer Einheitsmatrix mit Einsen in der Hauptdiagonale und Nullen in der oberen und unteren Dreiecksmatrix entspricht. In solch einem Modell mit zwei Klassen befinden sich 1.277 oder 73,51% in der ersten Klasse und 460 oder 26,49% in der zweiten Klasse. Die erste weist die höheren Wahrscheinlichkeiten einer positiven Zufriedenheitsantwort auf. Das Modell zeigt einen *loglikelihood*-Wert von -7.574,864, einen BIC Wert von 15.231,787 und einen ABIC von 15.196,841 mit elf frei geschätzten Parametern. Ein LRT-Vergleich mit dem stationären Modell weist eine Devianz von 149,116 mit zwei Freiheitsgraden und

einen resultierenden Signifikanzwert von 0,0000 auf. Der Vergleich mit dem nicht-stationären Modell weist eine Devianz von 167,144 mit zehn Freiheitsgraden und einer Signifikanz von 0,00000 auf. Somit wird von einer latenten Veränderung über die Zeit ausgegangen, welche in einem späteren Modell noch genauer spezifiziert wird.

4.3.7.38. Modellerweiterungen

Sonderfälle, wie beispielsweise ein *latent Markov plus random response* sind denkbar, wobei für eine Klasse angenommen wird, dass die Antworten zufällig sind, indem die Antwortwahrscheinlichkeiten auf 0,5 gesetzt werden, also *equiprobability*. Langeheine und Van de Pol (2002) zeigen auch die Erweiterung des LM-Modells mit einer Kette auf eines mit multiplen Ketten, das sogenannte LMM-Modell. Dabei kommt es leicht zu Identifikationsproblemen. Das Modell mit zwei latenten Ketten, nämlich ein *latent mover-mover*-Modell, und das mit einer latenten *mover*-Kette und einer latenten *stayer*-Kette, wobei nicht perfekte Messung für beide Ketten in beiden Modellen angenommen wird, sind für konstante Übergangswahrscheinlichkeiten identifizierbar. Sie sind jedoch beide schwach und weisen hohe Standardfehler in den geschätzten Parametern auf. Eine Verbesserung des Fits gegenüber dem LM-Modell mit einer Kette ist meist beträchtlich. Werden die Übergangswahrscheinlichkeiten als zeitlich heterogen spezifiziert, sind beide Modelle nicht identifizierbar. Dies heißt jedoch nicht, dass LM-Modelle mit zwei Ketten für Fünf-Wegedaten nicht brauchbar sind. Das *partially latent mover-stayer*-Modell hat eine latente *mover*-Kette mit und eine manifeste *stayer*-Kette ohne Messfehler. Deshalb wird es auch *partially latent* genannt.

Das MM-, als auch das LMM-Modell, erlauben die Aufdeckung nicht beobachteter Heterogenität. Die Stichprobe wird aufgrund der latenten nicht beobachteten quantifizierten Kettenanteile aufgesplittet. Die optimale Teilung minimiert die Unterschiede zwischen beobachteten und erwarteten Häufigkeiten, wobei jedoch nichts über die Bedeutung der latenten Variable verraten wird. Es kann lediglich auf unterschiedliche Dynamik in den Teilstichproben geschlossen werden. Durch die Miteinbeziehung von externen kategorialen Variablen im LMM-Modell für mehrere Gruppen, wird den Gruppen eine unterschiedliche Dynamik unterstellt. Es ist ratsam, anstatt der separaten Analyse für jede der Subgruppen, eine simultane Analyse durchzuführen. Das Modell erlaubt daher beobachtete, als auch nicht beobachtete Heterogenität. Zwei extreme Spezifizierungen sind denkbar. Entweder jede Gruppe wird ihren eigenen Parameterschätzungen überlassen. Dies kommt der getrennten Analyse der Subgruppen gleich, oder alle Sets an Parametern werden über alle Subgruppen

gleichgesetzt, was vollständige Homogenität bedeutet. Zwischen diesen Extremen gibt es eine Vielzahl an definierbaren Möglichkeiten für partielle Homogenität, wo für alle oder einige Subgruppen Sets an Parametern gleichgesetzt werden. Spezielle *within-group*-Restriktionen können spezifiziert werden. Theoretische Überlegungen sollten dabei die Vielzahl eingrenzen. Es muss bedacht werden, dass die Stichprobengröße durch die Gruppierung sinkt und die Parameterschätzungen an Exaktheit einbüßen und höheren Standardfehlern unterliegen.

Langeheine und Van de Pol (2002) stellen in genanntem Artikel lediglich Modelle mit einem Indikator vor, jedoch auch Modelle mit mehreren Indikatoren sind denkbar, was einem LC-Modell zu jedem Zeitpunkt gleichkommt. Vorteile liegen darin, dass die bedingten Antwortwahrscheinlichkeiten zu Identifikationszwecken nicht als zeithomogen begrenzt werden müssen. Eine Interpretation wird jedoch schwierig, wenn Veränderungen der Indikatoren erlaubt werden, da sich dann die Bedeutung der latenten Variable ändert. Ein Problem von multiplen Indikatoren liegt in der Spärlichkeit der Zellen, da die Kreuztabellen sehr groß werden. Zusätzlich zu den *time-constant* exogenen Kovariaten können auch *time-varying* Kovariate eingeführt werden. Die latenten Zustände werden dann sowohl auf die *time-constant*, als auch auf die *time-varying* Kovariaten in einem System an *logit*-Regressionen regressiert, was der Erweiterung des *Goodman causal loglinear*-Modells für beobachtete Zustände auf latentem Level gleichkommt. Dieses Modell ist sparsamer, verglichen mit dem *multiple-group*-Modell. Es kann jedoch wieder zu Spärlichkeitsproblem führen. Auch Modelle höherer Ordnung wie *second-order* sind denkbar. Diese haben jedoch einen schlechten Ruf, da sie durch die erhöhte Anzahl an Parametern, verglichen mit Modellen erster Ordnung zu einem besseren Fit führen. In der Wirklichkeit können jedoch nicht immer Abhängigkeiten höherer Ordnung gerechtfertigt werden. Deshalb sollten Modelle höherer Ordnung unbedingt theoretisch fundiert werden. Spärlichkeit führt nicht notwendigerweise zu Schätzproblemen. Die Evaluierung der Modelle für χ^2 -basierte Statistiken ist jedoch verletzt, da die wahre Verteilung dieser Statistiken schlecht an die theoretische χ^2 -Verteilung approximiert werden kann. Deshalb wird BIC verwendet. Eine weitere Lösung bietet die Simulierung der Referenzverteilung des verwendeten Fitmaßes für die Modell-Daten Kombination durch den Gebrauch von *bootstrap*-Methoden.

Vermunt (2004) erwähnt, dass die Erweiterungen des Markov *chain*-Modells hin zum *mover-stayer*-Modell im biomedizinischen Feld auch im Falle der *long-term survivors* angewendet werden, oder in der Marketingforschung auch als *brand-*

loyal bzw. *brand-switching segments*. Erweiterungen wie das *independence-stayer*-Modell, wobei die *mover* zufällig handeln, oder absorbierende Stadien, wie beispielsweise der Tod, wo ein Übergang zurück nicht möglich ist, sind ebenfalls denkbar.

Kaplan (2008) merkt an, dass es Markov-Modelle auch schon im *multilevel framework* gibt. Dadurch können latente Zustandsvariablen mit *random intercepts*, welche über Gruppen variieren, modelliert werden und diese Variabilität als Funktion von *between-groups* Kovariaten beschrieben werden. *Multilevel mover-stayer*-Modelle können aufzeigen, ob es *between-group*-Variabilität auf der höheren Ebene im Anteil von *movers* und *stayers* gibt und wie *between-group*-Kovariate diese Unterschiede erklären.

4.3.7.39. Die latente Kovariate in Form einer latenten Klassenvariable

Mc Donough et al. (2009) verwenden *mixture latent Markov* (MLM) Modelle um den Gesundheitsstatus einer Bevölkerung über die Zeit hinweg zu analysieren. Dafür wird eine latente Klassenvariable zur Analyse der Paneldaten mit einer kategorialen abhängigen *outcome*-Variablen verwendet. Es werden latente Gesundheitsstadien unterstellt, zwischen denen Übergänge über die Zeit hinweg auftreten und zusätzlich latente Klassen von Personen, welche unterschiedliche Übergangswahrscheinlichkeiten zwischen diesen latenten Klassen über die Zeit zeigen. Im Messmodell werden die latenten Gesundheitszustände über LCA geschätzt und im Strukturmodell die latenten gesundheitlichen Veränderungsprozesse. Die beobachteten Daten werden als Indikatoren eines wahren nicht beobachteten Gesundheitsstatus mit Fehler angesehen, wobei dieser als gleich über die beiden Populationen, USA und GB, angenommen wird. Der systematische *bias* der unterschiedlichen Items wird in den LM-Modellen durch Erlaubnis unterschiedlich beobachteter Antwortverteilungen, gegeben den gleichen latenten Gesundheitsstatus in den beiden Ländern, gewährt. In einem *mover-stayer*-Modell werden drei latente Gesundheitsprozesse aufgefunden, eine *mover*-Trajektorie, deren Personen zwischen den latenten Stadien über die Zeit hinweg wandeln, eine konstant niedrige Gesundheitstrajektorie und eine konstant gute Gesundheitstrajektorie. Die Veränderungen der *mover* werden über die Zeit gleich gehalten, die der *stayer* auf null fixiert. Die erwarteten Häufigkeiten der Personen in der Übergangsmatrix werden erklärt durch die beobachteten Antworten über die Zeit, die zeitkonstante latente *mover-stayer*-Klasse *c*, zwei zeitvariierende latente Gesundheitsstadien, ein Vektor von zeitkonstanten sozioökonomischen Prädiktoren und Kontrollvariablen zum Beobachtungszeitpunkt null, die Wahrscheinlichkeit der Zugehörigkeit zur latenten *mover-stayer*-Klasse *c* gegeben die Hintergrundprädiktoren, die Verteilung der latenten

Gesundheitsstadien zum Zeitpunkt t in der *mover*-Klasse c , die Übergangswahrscheinlichkeit zwischen zwei latenten Gesundheitsstadien in der *mover*-Klasse c welche über die Zeit konstant ist, und durch die bedingten Antwortwahrscheinlichkeiten in der *mover-stayer* Klasse c gegeben den latenten Status zu wandeln, wobei diese über alle Zeitpunkte und *mover-stayer* Klassen gleich sind. Geschätzt wurde mit FIML unter MAR Bedingung.

Die Spezifizierung einer latenten Kovariate ist in Form einer latenten Klassenvariable höherer Ordnung möglich, wobei die latente Variable höherer Ordnung auf die latenten Klassenvariablen zu jedem Zeitpunkt bezogen wird, wie in Diagramm 145 dargestellt.

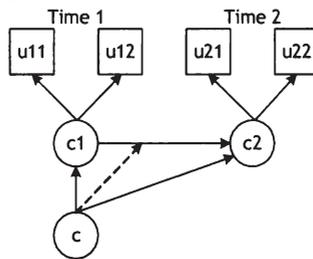


Diagramm 145: Second order latent class covariate (Muthén und Muthén, 2007)

Der latenten Klassenvariablen wird eine vorweg angenommene Anzahl an Klassen unterstellt. Wechselwahrscheinlichkeitsrestriktionen können für jede Klasse der latenten Klassenvariablen höherer Ordnung eingeführt werden, um spezielle Klassen von Individuen basierend auf Übergängen zu identifizieren. Ein Spezialfall der latenten Klassenkovariate höherer Ordnung ist die Spezifizierung von *movers* und *stayers*. Die latente Klassenvariable höherer Ordnung, in diesem Fall die *mover-stayer*-Variable, wird verwendet, um die nichtbeobachtete Heterogenität in den Wechselwahrscheinlichkeiten aufzufangen. Die *mover* wechseln zwischen den Klassen und die *stayer* verbleiben in der gleichen Klasse über die Zeit. Die *mover-stayer*-Variable differenziert also Personen, welche in derselben Klasse über die Zeit verbleiben, von jenen, welche zumindest einen Wechsel verzeichnen. Durch die Trennung der *mover* und *stayer* werden die Transitionswahrscheinlichkeiten der *mover* genauer geschätzt, vorausgesetzt es gibt tatsächlich Personen mit einer Wahrscheinlichkeit von null zwischen den Klassen zu wechseln. Die Übergangsmatrix der *mover* ist nicht beschränkt und die Übergangswahrscheinlichkeiten werden frei geschätzt. Die der *stayer* ist dahingehend beschränkt, dass die Hauptdiagonalwerte auf eins fixiert werden und alle Nebendiagonalwerte auf null. Die Übergangsmatrix nimmt somit die Gestalt

einer Identitätsmatrix an. Dieses Modell ist nur sinnvoll, wenn Klassenanzahl und -typus gleich bleiben. Ist *full measurement invariance* nicht gewährleistet, macht die Interpretation der *stayer* keinen Sinn, da sich die Klasseneigenschaften über die Zeit hinweg verändern können.

Es wird hier davon ausgegangen, dass eine Subpopulation existiert, welche keine Übergänge über die Zeit hinweg aufweist, und eine andere welche Übergänge zu verzeichnen hat. Dieses Modell wird *mover-stayer*-Modell genannt und fällt in die Klasse der MLM-Modelle. Die Stichprobe wird folglich segmentiert und das *full* LTA-Modell mit einer zusätzlichen latenten Klassenvariable geschätzt. Die *stayer* haben eine Wechselwahrscheinlichkeit von null. Als Output erhält man die Klassenverteilungen der *mover* und *stayer*, aufgeteilt auf die Klassen und die Übergangswahrscheinlichkeitsmatrizen. Da die *stayer* alle in Klasse zwei sind, ist es nicht notwendig, für die *mover* separate Übergangswahrscheinlichkeiten zu berechnen, da die *stayer* die Übergangswahrscheinlichkeiten der *mover* nicht beeinträchtigen. Es zeigen sich lediglich kleine Unterschiede in den Übergangswahrscheinlichkeiten. Ein Vergleich zwischen den Übergangswahrscheinlichkeiten mit der LTA mittels BIC zeigt bei Kaplan (2008) die Bevorzugung des *mover-stayer*-Modells, also einen besseren Fit der manifestierten Antworthäufigkeiten verglichen mit der LTA. Es zeigen sich ebenfalls Unterschiede zwischen LR- und Pearson- χ^2 , was auf schwach besetzte Zellen hindeutet und einer detaillierten Untersuchung bedarf.

4.3.7.40. Mover-stayer model - Kinodatensatz

Im Kinobeispiel gibt es annähernd die gleichen Klassen und Klassentypen über die Zeit hinweg. Somit macht die Einführung einer *mover-stayer*-Variablen durch Spezifikation einer latenten Variablen höherer Ordnung Sinn. Es wird angenommen, dass es zu jedem Zeitpunkt gleiche Anzahl und gleiche Typen an Klassen gibt. Im später gezeigten *latent mixed* Markov-Modell, genauer gesagt in der Spezifizierung eines *mover-stayer*-Modells, werden keine Kovariate miteinbezogen und es wird von einer stationären Wechselwahrscheinlichkeit mit *first-order* Effekt ausgegangen. Das Modell teilt die Stichprobe in zwei Gruppen, die *mover* und die *stayer*, deren Größen ausgegeben werden. Für den Kinodatensatz wird unterstellt, dass es eine *stayer*-Gruppe gibt, welche über die Zeit hinweg in derselben Klasse eins oder zwei verbleibt. Weiters wird von einer *mover*-Klasse ausgegangen. Personen dieser Subpopulation dürfen über die Zeit variieren. Diese Spezifikation führt zu einer genaueren Schätzung der Ergebnisse, da jene Personen, welche einen stabilen Verlauf über die Zeit hinweg aufweisen, also die eigentlichen *stayer*, in die Berechnung der Übergangswahrscheinlichkeiten der *mover* nicht mehr eingehen. Um die beiden

Gruppen zu beschreiben, können einerseits die sich verändernden Gruppen-
größen aller *mover*- und *stayer*-Klassen gesamt betrachtet werden. Andererseits
können zwecks separater Beschreibung der *mover* und *stayer* die am häufigsten
auftretenden Patterns betrachtet werden. Die Gruppe der *stayer* wird durch ihre
%-uelle Aufspaltung in die einzelnen Gruppen, welche über die Zeit hinweg
gleich bleiben, beschrieben. Also bei drei Klassen, drei Prozentwerte der
Patternhäufigkeit für die drei Möglichkeiten der gleichbleibenden Klasse. Die
Gruppe der *mover* wird aufgrund unterschiedlicher Veränderungen
unterschiedliche Patterns aufweisen. Die am häufigsten auftretenden Patterns
dieser Gruppe stehen für die allgemeine Veränderung der Personen. Nylund
(2007) verwendet dafür die Patterns mit einer Häufigkeit über zehn. Die Wahl
der Patterngröße hängt natürlich von der Stichprobe ab. Um das Problem
spärlicher Zellen zu vermeiden und die Pearson- χ^2 -*statistic* und LR- χ^2 -*statistic*
zur Modellfittestung verwenden zu können, werden die Daten wiederum
dichotomisiert.

Diagramm 146 zeigt den *item probability plot* für das *mover-stayer*-Modell. Die
latenten Zustände sind über die *mover* und *stayer* gleichgesetzt. Es kristallisiert
sich eine zufriedene Klasse zwei heraus und eine weniger zufriedene Klasse eins.
Beide Klassen stellen das Messmodell für das *mover-stayer*-Modell dar, also
sowohl für die *mover*, als auch für die *stayer*. Das Modell zeigt einen
loglikelihood-Wert von -7.376,506, einen BIC von 14.924,590 und ABIC von
14.851,521.

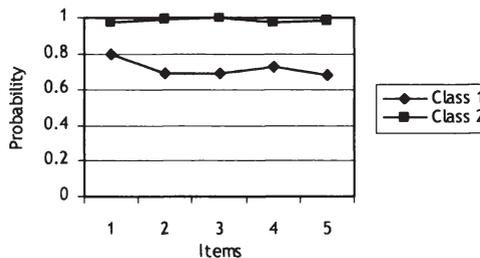


Diagramm 146: Mover-stayer item probability plot

In Tabelle 99 sind die endgültigen Klassenhäufigkeiten und Anteile der latenten
Klassenpatterns, basierend auf dem geschätzten Modell, gegeben. Im oberen Teil
befinden sich die Klassenpatterns, Häufigkeiten und Anteile der *mover*, im
unteren Teil die der *stayer*.

		Latent class pattern over six month						Respon- dents	Per- cent
		Month 1	Month 3	Month 5	Month 7	Month 8	Month 11		
Movers 687.62 39.59%	2	2	2	2	2	2	94.42	5,44	
	2	2	1	2	2	2	50.15	2,89	
	1	2	2	2	2	2	36.75	2,12	
	2	2	2	2	2	1	31.44	1,81	
	2	1	1	2	2	2	29.90	1,72	
	2	1	2	2	2	2	29.81	1,72	
	2	2	2	2	1	1	23.21	1,34	
	2	2	2	1	2	2	20.87	1,20	
	2	2	1	1	2	2	20.10	1,16	
	1	2	1	2	2	2	19.52	1,12	
	2	2	2	2	1	2	18.71	1,08	
	2	2	1	2	2	1	16.70	0,96	
	2	2	1	2	1	1	12.33	0,71	
	1	2	2	2	2	1	12.24	0,70	
	2	1	1	1	2	2	11.98	0,69	
	2	2	2	1	1	1	11.92	0,69	
2	2	1	1	1	1	11.48	0,66		
Stayers 1.049.38 60.41%	1	1	1	1	1	1	42.50	2.45	
	2	2	2	2	2	2	1,006.87	57.9	

Tabelle 99: Antwortpatternhäufigkeiten und –anteile des mover-stayer Modells

1.006,87 der Respondenten befinden sich in Klasse zwei der *stayer*, weisen also eine hohe Wahrscheinlichkeit einer zufriedenen Antwort über die Zeit hinweg auf. 42,5 Personen weisen eine niedrige Wahrscheinlichkeit einer zufriedenen Antwort auf. Das Bild der *mover* zeigt ein gemischtes Bild. Das am häufigsten auftretende latente Klassenpattern, nämlich 2 2 2 2 2 2, ist ident mit dem am häufigsten auftretenden Pattern der *stayer*, nämlich über die Zeit hinweg in Klasse zwei zu verbleiben. An zweiter Stelle mit 2,89% aller Antworten und etwa 50 Probanden findet sich das latente Klassenpattern 2 2 1 2 2 2 und zeigt ebenfalls sehr hohe Zufriedenheit über die Zeit.

Um abschließend einen leichteren Überblick zu bekommen, werden in Tabelle 100 die Häufigkeitszahlen der beiden latenten Zustände für die *stayer* und *mover* aufsummiert. Es zeigt sich eine sinkende Wahrscheinlichkeit für positive

Zufriedenheitsantworten, da mehr Personen aus der latenten Klasse zwei, über die Zeit in die latente Klasse eins wandern.

Time point	1	3	5	7	9	11
Class one	226	239	339	234	250	277
Class two	1,511	1,498	1,398	1,503	1,487	1,460

Tabelle 100: Absolute Häufigkeiten der latenten Zustände im mover-stayer Modell

Der sinkende Zufriedenheitstrend taucht in zwei Teilen auf, nämlich von Zeitpunkt eins bis fünf und wieder von Zeitpunkt sieben bis elf. Der sinkende Trend ist nicht konstant über sämtliche Wellen. Derartige Verläufe zwischen latenten Zuständen unter Miteinbeziehung unterschiedlicher Subpopulationen, welche unterschiedlichen Verteilungen folgen, können nur mit Hilfe nicht-stationärer Übergangswahrscheinlichkeiten diskreter Zeitmodelle aufgedeckt werden. Hauptproblem dieser Analysen liegt im enormen Rechenaufwand der Modelle. Deshalb wurde dieses abschließende Beispiel auch lediglich mit der Dateninformation aus sechs Beobachtungen und zwei Subpopulationen mit jeweils nur zwei Klassen geschätzt. Zusammengefasst liefert diese Modellgruppe jedoch eine enorme Vielfalt an Spezifizierungsmöglichkeiten, welche durch effizientere Schätzer oder stärkere Hardware in naher Zukunft gelöst werden sollten.

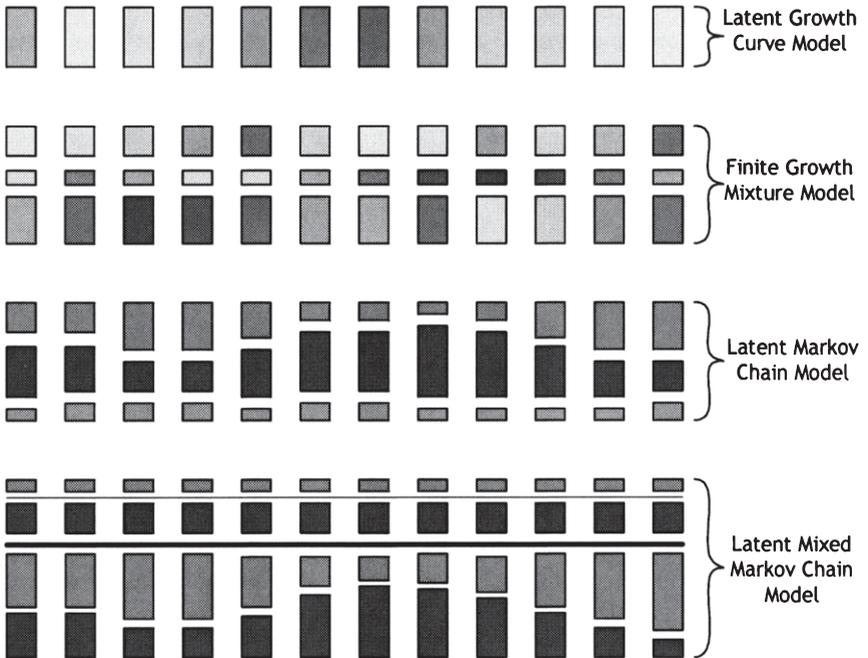
5. Schlussfolgerungen und Verwertbarkeit

Der Großteil der in der Literatur vorhandenen Zufriedenheitsstudien untersucht den Einfluss einzelner Zufriedenheitsindikatoren auf die Gesamtzufriedenheit. Diese Items werden auf Basis der Ergebnisse unterschiedlicher Methoden in Faktoren klassifiziert, welche im Wesentlichen aus dem ursprünglichen Kano-Modell abgeleitet sind. Nachteile liegen in der meist nicht beachteten Annahme, dass aufgrund von Störfaktoren die Messung fehlerbehaftet ist. Von einer perfekten Messung kann jedoch nicht ausgegangen werden kann. So wird die wahre Varianz mit anderen Varianzquellen vermischt und führt zu verzerrten Ergebnissen. Einige Studien wurden auf Basis latenter Konstrukte zwecks Steigerung der Reliabilität und Validität durchgeführt. Es sind aber lediglich *cross-sectional*-Studien auffindbar. Ein gesteigertes Auftreten von Zufriedenheitsuntersuchungen auf latenter Ebene ist in der jüngeren Vergangenheit bemerkbar, da eben eine perfekte Messung beobachteter Indikatoren speziell in sozialempirischen Befragungen psychologischer Konstrukte nicht gewährleistet werden kann.

Die meisten Studien basieren, wie gerade erwähnt, auf *cross-sectional*-Daten. Es wurde bereits oftmals postuliert und als Forschungsempfehlung appelliert, dass zeitabhängige Aspekte ebenfalls berücksichtigt werden müssen. Werden diese vernachlässigt, wird zu jedem einzelnen Zeitpunkt nur eine Momentaufnahme analysiert. Hinsichtlich der Überlegung, dass sich auch die Messmodelle ändern können, führt dies speziell auch im Bereich der Erstellung von Messmodellkonstrukten zu Fehlinterpretationen. Deshalb ist es notwendig, zeitabhängige und zeitunabhängige Aspekte eines Modells simultan zu untersuchen. Doch nicht nur alleine aufgrund von zeitlichen Komponenten bilden *cross-sectional*-Modelle die Situationen realitätsfremd ab.

Ein weiterer Aspekt liegt in dem Umstand, dass in der besprochenen Zufriedenheitsliteratur beinahe immer nur von einer Population ausgegangen wird. In der Realität wird sich dies nur schwer bewahrheiten. Loyalitätsgruppierungen am Kinodatensatz zeigten zufriedenheitsspezifische Unterschiede auch auf loyalitätssegmentspezifischem Level. Je nach Loyalitätsstatus wirken unterschiedliche Einflüsse auf verschiedene Weise in abweichender Stärke. Deshalb müssen Mischungen in die Modelle eingebracht werden. Größenmäßig unterschiedliche segmentspezifische Entwicklungen würden die Messmodelle nämlich ebenfalls verzerren.

Ein weiterer Punkt liegt in der meist falschen Verwendung ordinal skaliertter Fragen. Analysen werden aus jenem Gesichtspunkt ausgeführt, als ob die Items intervallskaliert wären. Die innewohnenden Probleme dieser falschen Betrachtungsweise werden in der vorliegenden Arbeit ausführlich besprochen und dürfen nicht außer Acht gelassen werden. Diagramm 147 stellt einen visualisierten Überblick zu den aufgezählten Problemfeldern dar.



Alle Modelle aus der Graphik lösen das Problem der Messfehler über die Einführung einer latenten Zwischenebene im Messmodell. In oberen beiden ist dies ein kontinuierlicher Faktor, in den unteren beiden ein diskreter Faktor. Somit scheint das erste Problemfeld ungenauer Beobachtung gelöst.

Zweites Problemfeld liegt in der unterschiedlichen Entwicklung von Subgruppen. *Latent growth curve*-Modelle beschreiben lediglich eine einzige Population, welche sich über die Zeit verändert. *Finite growth mixture*-Modelle beinhalten zusätzlich die Möglichkeit, mehrere Populationen über die Zeit hinweg zu untersuchen. Sie inkludieren somit den erwähnten Segmentierungsansatz. Das Messmodell entspricht der latenten Darstellung der manifesten

Indikatoren auf einem Kontinuum. Das Strukturmodell ermöglicht die latente Veränderung auf einem höherliegenden Level und eine zusätzliche diskrete Einteilung in Subpopulationen.

Verglichen mit den beiden oben genannten Modellen weisen *latent Markov chain*-Modelle die Möglichkeit auf, Personen zwischen den latenten Klassen Sprünge über ein Zeitintervall zu erlauben. Die latenten Klassen selbst verändern sich nicht. Somit stellt das Messmodell eine gleichbleibende latente Eigenschaftsrepräsentation auf einem diskreten Level dar. Das Strukturmodell zur Subpopulationsunterscheidung ist ebenfalls diskret. Das *latent mixed Markov chain*-Modell ermöglicht nämlich zusätzlich die simultane Segmentierung in Subpopulationen mit unterschiedlichen Sprungeigenschaften der Subjektgruppen.

Die innewohnenden Eigenschaften konstanter Messmodelle stellen aus schätz- und interpretationstechnischen Gründen meist eingeführte Bedingungen dar, welche aber nicht zwangsbedingt notwendig sind. Sie vereinfachen aber deutlich die Modellierungsarbeit, Präferenzierungs- und Interpretationsverfahren. Zur Auflösung genannter Bedingungen in beiden Modellklassen stehen einige Möglichkeiten zur Verfügung. Beispielsweise kann in beiden Modellklassen zeitliche Variation über das Messmodell ermöglicht werden, in den Wachstumsmodellen über freie Faktorenladungsstrukturen und in den Klassenmodellen über frei variierende Klassenprofile. Dies bringt eine große Vielfalt an Variationsmöglichkeiten in die Modelle ein. Eine Interpretation sich verändernder kontinuierlicher oder diskreter Dimensionen erscheint jedoch kaum möglich und zusätzlich steigt die Anzahl der freigeschätzten Parameter enorm an.

Auch eine Variation der Subpopulationsanzahl über die Zeit ist denkbar. Klassen lösen sich auf oder werden im zeitlichen Fluss neu geboren. Die aktuellsten hier zuletzt durchgeführten Modellspezifikationen erlauben es zumindest, Subpopulationen unterschiedlicher Verteilungen in einem Modell mit unterschiedlichen Entwicklungen zu schätzen. Auch Modelle, welche die Schätzung entstehender oder sterbender latenter Klassen ermöglicht, wurden bereits entwickelt und hier auch kurz angerissen.

Die im letzten Kapitel aufgetauchte Diskussion über Markovianische Strukturen unterschiedlicher Ordnungen eröffnet zusätzlich einen, im Zufriedenheitsbereich bis dato nicht diskutierten, enormen Spielraum an Variationsmöglichkeiten. Eine Zusammenführung sämtlicher Modellspezifikationen erscheint aus schätztechnischer Sicht noch nicht effizient genug. Die Vielfalt an Möglichkeiten, welche in

den letzten Absätzen nur kurz angerissen wurde, ist über die verfügbaren Modellselektionsmöglichkeiten bis dato kaum bis gar nicht bewältigbar.

Abschließend wird hier nochmals auf das letzte Diagramm hingewiesen und eine interessante innewohnende Kontroverse diskutiert. Die oberen beiden, verglichen mit den unteren beiden Modellklassen, weisen nämlich einen markanten Unterschied auf, welcher bis dato aufgrund fehlender Studien noch nicht diskutiert wurde. In den ersten beiden Modellen verbleiben die Personen über den gesamten Zeitraum in derselben Population. Sie springen nicht in andere Populationen mit unterschiedlichen Entwicklungen bzw. Eigenschaften. Die Klassen verändern nämlich selbst über die Zeit ihre Eigenschaften. In den unteren beiden Modellen wird den Personen ermöglicht, zwischen den Klassen zu wechseln. Die unterschiedlichen Klassen verändern sich jedoch nicht, lediglich die Klassenanteile.

Die wahrscheinlich derzeit modellspezifisch beste Übersetzung der Realität liegt in den Augen des Autors in den *latent mixed Markov chain*-Modellen, und über die Erweiterung der Einführung zeitlich variierender latenter Profile, Klassen- und Subpopulationsanzahl, jegliche Wandlungsweise zu erlauben und durch wohl überlegtes Herunterbrechen den Interpretationsspielraum zu schmälern. Bis zur Entwicklung durchführbarer und vor allem auch effizienter Schätzungs-lösungen kann hier nur auf Basis der angesprochenen Punkte die grundsätzliche Diskussion über eine generelle Herangehensweise für den Analysebeginn im Bereich der Zufriedenheitsentwicklungsformen eröffnet werden.

Derzeit stellt sich wohl aus verwirklichungstechnischen Gründen zuallererst die Frage, mit welcher übergeordneten Modellklasse begonnen wird. Entweder man legt den Analyseschwerpunkt in variablen- oder personenzentrierte Hände, also

*in die Existenz von Subpopulationen, deren Subjekte homogenen Veränderungen unterliegen
und sich in einer endlichen Anzahl an Klassen über die Zeit hinweg wandeln,
oder in die Existenz endlicher gleichbleibender Subpopulationseigenschaften,
zwischen welchen die zeitliche Wandlung der Individuen von statten geht.*

6. Diagrammverzeichnis

Diagramm 1: Action Grid (Crompton und Duray, 1985)	25
Diagramm 2: Slack's IPA Matrix (Sampson und Showalter, 1999)	27
Diagramm 3: Importance Performance Zusammenhang (Sampson und Showalter, 1999)	29
Diagramm 4: One-dimensional vs. two-dimensional recognition method (Kano et al., 1984)	31
Diagramm 5: Satisfaction/dissatisfaction Kriterien anhand der Critical Incident Technique (Johnston, 1995)	32
Diagramm 6: Kano's model of customer satisfaction (Matzler et al., 1996)	34
Diagramm 7: Asymmetric impact of attribute-level performance (Mittal et al., 1998)	38
Diagramm 8: Performance-Satisfaction Link (Anderson und Mittal, 1998)	40
Diagramm 9: Prospect Theory (Audrian-Pontevia, 2006)	44
Diagramm 10: Kohonen Self-Organizing Map mit Satisfaction Maintainer (Audrian-Pontevia, 2006)	45
Diagramm 11: Penalty-reward contrast-Analyse (Matzler und Sauerwein, 2002)	49
Diagramm 12: Wichtigkeit als abhängige Variable von Performance (Matzler, 2004)	53
Diagramm 13: Gruppenvergleich der IPA matrix (Matzler, 2004)	53
Diagramm 14: Kano Modell (Vargo et al., 2007)	55
Diagramm 15: Monovalent vs. bivalent (Vargo et al., 2007)	55
Diagramm 16: Zone of Indifference (Vargo et al., 2007)	56
Diagramm 17: Kano-Modell übertragen auf die Kundenzufriedenheit (Buhl et al., 2007)	58
Diagramm 18: Modifiziertes Kano-Modell (Buhl et al., 2007)	59
Diagramm 19: Dimensionen der Wichtigkeit (Van Ittersum et al., 2007)	61
Diagramm 20: Zufriedenheitskonstrukte (Campo und Yagüe, 2009)	64
Diagramm 21: From Satisfaction to delight (Toores und Kline, 2006)	66
Diagramm 22: Delight and Satisfaction (Finn, 2005)	67
Diagramm 23: Revised model (Ladhari, 2007)	69
Diagramm 24: Ordered satisfaction clusters (Mazanec, 2006)	71
Diagramm 25: Contributions to Overall Satisfaction (Mazanec, 2006)	72
Diagramm 26: Wirkungsbeiträge von vier ausgewählten Zufriedenheitsfaktoren (Mazanec, 2007)	73
Diagramm 27: Zone-of-tolerance (Pollack, 2008)	74
Diagramm 28: Acceptable level (Pollack, 2008)	75
Diagramm 29: Faktorscores (Klein und Muthén, 2007)	76

Diagramm 30: Importance Performance Grid	91
Diagramm 31: Dummy-Regression 12345-6	94
Diagramm 32: Dummy-Regression 1234-56	96
Diagramm 33: Vavras-Importance-Grid (Welle zwei)	99
Diagramm 34: Vavras-Importance-Grid (Welle zwölf)	101
Diagramm 35: Zufriedenheit und Wichtigkeit der Lieblingskinobesucher	105
Diagramm 36: Zufriedenheit und Wichtigkeit der Nicht-Lieblingskinobesucher	106
Diagramm 37: Zufriedenheit und Wichtigkeit	109
Diagramm 38: Nutzenwerte (beide Zeitpunkte)	119
Diagramm 39: Paarvergleichswichtigkeiten im zeitabhängigen Vergleich	120
Diagramm 40: Modell Hierarchie (Mair und Hatzinger, 2007)	138
Diagramm 41: Designmatrix wiederholter Messungen mit Subgruppen (Mair und Hatzinger, 2007)	139
Diagramm 42: Bestimmung polytomer IRT-Modelle (Ostini und Nering, 2006)	142
Diagramm 43: Joint-Plot der ICC-Kurven - dichotomes Modell (123-456)	147
Diagramm 44: Joint-Plot der ICC-Kurven - dichotomes Modell (12-3456)	148
Diagramm 45: Graphischer Modelltest (123-456)	150
Diagramm 46: Standort vs. Buffet im RSM	153
Diagramm 47: Threshold Probability Curves	154
Diagramm 48: Wahrscheinlichkeitsübergänge über die ICCs	155
Diagramm 49: ICCs	156
Diagramm 50: Person-Item Location Verteilung - Zufriedenheit	156
Diagramm 51: Person-Item Threshold Verteilung	157
Diagramm 52: Thresholds des PCM	158
Diagramm 53: Locations des PCM	159
Diagramm 54: ICC Plots des PCM	160
Diagramm 55: ICC-Vergleich zweier Standorte für das Item Standort/Erreichbarkeit	161
Diagramm 56: Person-Item Location Verteilung - Wichtigkeit (November 2006)	162
Diagramm 57: Item Thresholds und Locations – Wichtigkeiten (November 2006)	162
Diagramm 58: IPA mit Zufriedenheits- und Wichtigkeits-Locations	163
Diagramm 59: Zeitliche Gruppenlösung – Wichtigkeiten	164
Diagramm 60: LLRA - Zufriedenheit	166
Diagramm 61: LLRT-Wichtigkeit	168
Diagramm 62: Object Plot	174
Diagramm 63: Joint Plot	175

Diagramm 64: Span Plot	176
Diagramm 65: Ladungsplot	176
Diagramm 66: Graphplot	177
Diagramm 67: Starplot - Filmangebot	178
Diagramm 68: Voronoi Plot - Gesamtzufriedenheit	179
Diagramm 69: Voronoi Plot - Film	179
Diagramm 70: Kategorienplot - nicht-monoton – Technik/Bild/Ton	179
Diagramm 71: Hullplot – Labplot - Gesamtzufriedenheit	180
Diagramm 72: Lossplot - Buffet	181
Diagramm 73: Joint Plot Vergleich	181
Diagramm 74: Joint Plot Vergleich - Preis/Leistung	182
Diagramm 75: Kategorienplot - Gesamtzufriedenheit	182
Diagramm 76: Ladungsplot - lineare vs. monotone Regression	183
Diagramm 77: Vergleich der originalen und transformierten Skala – linear vs. monoton	183
Diagramm 78: Vergleich - lineare vs. monotone Regression	184
Diagramm 79: Labelplot - Komfort	184
Diagramm 80: Vektorplot und Projektionsplot	185
Diagramm 81: Random Intercept-Modell - Technik/Bild/Ton	196
Diagramm 82: Random Slope Modell – Technik/Bild/Ton	198
Diagramm 83: Nicht-lineares Random Slope Modell – Technik/Bild/Ton	200
Diagramm 84: Random Intercept-Modell - Residuen und Konfidenzintervalle	201
Diagramm 85: Standortvergleich bei gleichem Intercept	202
Diagramm 86: Standortvergleich bei gleichem Intercept und unterschiedlichen Slopes	203
Diagramm 87: Kundenkarteneinfluss - Technik/Bild/Ton	207
Diagramm 88: Level-2-Varianzveränderung - Technik/Bild/Ton	208
Diagramm 89: Level-1- und -2-Varianzveränderung - Technik/Bild/Ton	209
Diagramm 90: Varianzunterschiede Kundenkartenbesitzer vs. -nichtbesitzer	211
Diagramm 91: Kundenkartenbesitzer vs. -nichtbesitzer mit Slopevariation	213
Diagramm 92: Mittelwertverteilung - Kinostandorte	214
Diagramm 93: Level-zwei-Residuen und Konfidenzintervalle	220
Diagramm 94: Residuen	221
Diagramm 95: Variance Components Modell	222
Diagramm 96: Vorhergesagte Standortwerte für die Zufriedenheitsbewertungen	223
Diagramm 97: Vorhergesagte Werte für das Modell 26	225
Diagramm 98: Random Intercept Fixed Slope	226
Diagramm 99: Random Slope Random Intercept	226

Diagramm 100: Quadranteneinteilungen für Intercept und Slope Kombinationen	228
Diagramm 101: Konfidenzintervalle	228
Diagramm 102: Einfluss Cineplexx-Eigenschaft	232
Diagramm 103: Level-2-Varianzveränderung mit Technik/Bild/Ton- Zufriedenheit	234
Diagramm 104: Level-1- und -2-Varianz über Technik/Bild/Ton- Zufriedenheit	235
Diagramm 105: Varianzvisualisierung der Besuchertypen	236
Diagramm 106: Varianzvisualisierung für unterschiedliche Kinotypen	237
Diagramm 107: Antwortverteilung	238
Diagramm 108: Standardisierte Level-2-Residuen	245
Diagramm 109: Residuenverteilung	245
Diagramm 110: Zeitliche Varianzentwicklung	248
Diagramm 111: Individuelle Personenentwicklungen	248
Diagramm 112: Zeitliche Veränderung alle Probanden	249
Diagramm 113: Contour Plots longitudinale Zufriedenheitsbeobachtungen	249
Diagramm 114: Zeitlich individuelle Varianzen	250
Diagramm 115: Clustering über die Zufriedenheitsverläufe	252
Diagramm 116: Variation der Cluster	252
Diagramm 117: Lineares LGC-Modell	279
Diagramm 118: Quadratisches LGC-Modell	282
Diagramm 119: Frei geschätztes LGC-Modell	283
Diagramm 120: Piecewise LGC-Modell	285
Diagramm 121: Erweiterung zu getrennten Wachstumsverläufen	287
Diagramm 122: Linearer vs. nicht-linearer Verlauf	293
Diagramm 123: Linearer vs. nicht-linearer Verlauf mit Konfidenzintervallen	295
Diagramm 124: Modell mit indikatorspezifisch zeitübergreifenden diachronen Korrelationen	296
Diagramm 125: Intercept-Slope-Korrelationen	300
Diagramm 126: Drei-Klassenmodell latenter Wachstumskurven	318
Diagramm 127: General modeling framework (Muthén und Muthén, 2000)	325
Diagramm 128: Drei-Klassenlösung cross-sectional Vergleich	340
Diagramm 129: Hidden Markov Model (Muthén und Muthén, 2007)	341
Diagramm 130: Item probability plot - hidden Markov-Modell	342
Diagramm 131: Stage-sequential developmental-Modell (Collins et al., 2000)	343
Diagramm 132: LCA item probability plots geordneter (links) vs. nicht- geordneter (rechts) binärer Daten (Nylund, 2007)	348

Diagramm 133: Item probability plots	351
Diagramm 134: LTA – first-order Effekt (Nylund, 2007)	354
Diagramm 135: LTA - first- und second-order Effekt (Nylund, 2007)	354
Diagramm 136: Item probability plot	359
Diagramm 137: Item probability plot	361
Diagramm 138: Item probability plot Vergleich	362
Diagramm 139: Discrete latent Markov Modell mit Kovariaten (Vermunt et al., 1999)	362
Diagramm 140: Discrete latent Markov Modell mit Kovariaten und weiteren Effekten (Vermunt et al., 1999)	365
Diagramm 141: Mixed Markovmodell formuliert als latentes Markovmodell (Van de Pol und Langeheine, 1990)	379
Diagramm 142: Markovkettenmodellhierarchie (Langeheine und Van de Pol, 2002)	383
Diagramm 143: Classic Markov chain Modells (Langeheine und Van de Pol, 2002)	384
Diagramm 144: Latentes Markov Modell vs. latentes Klassenmodell (Langeheine und Van de Pol, 2002)	388
Diagramm 145: Second order latent class covariate (Muthén und Muthén, 2007)	392
Diagramm 146: Mover-stayer item probability plot	394
Diagramm 147: Modellvergleichsvisualisierung	398

7. Tabellenverzeichnis

Tabelle 1: Implications of hypotheses (Sampson und Showalter, 1999)	29
Tabelle 2: Functional/dysfunctional Form (Matzler et al., 1996)	35
Tabelle 3: Einteilung der Eigenschaftsausprägung (Matzler et al., 1996)	35
Tabelle 4: Fehleinschätzung des linearen symmetrischen Satisfaction- Performance Links I (Anderson und Mittal, 1998)	40
Tabelle 5: Fehleinschätzung des linearen symmetrischen Satisfaction- Performance Links II (Anderson und Mittal, 1998)	41
Tabelle 6: Action-steps chart (Anderson und Mittal, 1998)	43
Tabelle 7: Attribute – Overall Satisfaction Links (Audrian-Pontevia, 2006)	46
Tabelle 8: Importance Grid (Matzler und Sauerwein, 2002)	48
Tabelle 9: Importance-grid Empfehlungen (Matzler und Sauerwein, 2002)	50
Tabelle 10: Asymmetric Impact of Satisfaction Drivers (Füller et al., 2006)	52
Tabelle 11: Einteilung der Auslöser von Satisfaction und Dissatisfaction (Vargo et al., 2007)	54
Tabelle 12: Dummy-Regression 12345-6	93
Tabelle 13: Dummy-Regression 1234-56	95
Tabelle 14: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei)	97
Tabelle 15: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei, neu)	98
Tabelle 16: Explizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwei)	98
Tabelle 17: Implizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwölf)	100
Tabelle 18: Implizite Wichtigkeiten - Vavras-Importance-Grid Ergebnisse (Welle zwölf, neu)	100
Tabelle 19: Explizite Wichtigkeiten - Vavras-Importance-Grid (Welle zwölf)	101
Tabelle 20: Wichtigkeitsmittelwerte	107
Tabelle 21: Ranking der Wichtigkeitsmittelwerte	108
Tabelle 22: Wichtigkeitsparameter (Zeitpunkt eins)	113
Tabelle 23: Nutzenwerte (Zeitpunkt eins)	113
Tabelle 24: Wichtigkeitsparameter - Subjektkovariate (Zeitpunkt eins)	114
Tabelle 25: Nutzenwerte - Subjektkovariate (Zeitpunkt eins)	114
Tabelle 26: Wichtigkeitsparameter - Subjektkovariate - Interaktion undecided-Kategorie (Zeitpunkt eins)	115
Tabelle 27: Nutzenwerte – Subjektkovariate - Interaktion undecided- Kategorie (Zeitpunkt eins)	115
Tabelle 28: Modellvergleich der zeitabhängigen Modelle	117
Tabelle 29: Geschätzte Parameter des favorisierten zeitabhängigen Modells	118

Tabelle 30: Nutzenwerte (beide Zeitpunkte)	118
Tabelle 31: Klassische vs. probabilistische Testtheorie (Embretson und Reise, 2000)	124
Tabelle 32: Item Response Modell für die Veränderungsmessung (Glück und Spiel, 1997)	132
Tabelle 33: LLRA-Designmatrix	136
Tabelle 34: Locations - dichotomes Modell (123-456)	146
Tabelle 35: Locations - dichotomes Modell (12-3456)	149
Tabelle 36: Likelihood-Ratio-Test nach Andersen für das 123-456-Modell	149
Tabelle 37: Wald-Test (123-456)	151
Tabelle 38: Locations und Thresholds - RSM	152
Tabelle 39: Locations und Thresholds des PCM	158
Tabelle 40: Location- und Thresholdvergleich zweier Standorte für das Item Standort/Erreichbarkeit	160
Tabelle 41: Kontingenztafel - Wichtigkeit	167
Tabelle 42: Klassifizierungsrate	177
Tabelle 43: Prognosetabelle	178
Tabelle 44: Multilevel Modelle	188
Tabelle 45: Interceptvariation – Gesamt- und Teilzufriedenheiten	193
Tabelle 46: Random Slope-Modell - Teilzufriedenheiten	199
Tabelle 47: Standortmittelwerte und Abweichungen	216
Tabelle 48: Standortresiduen	222
Tabelle 49: Standardabweichungsvergleich	232
Tabelle 50: Anti-Logits	240
Tabelle 51: Kategorienprozentwerte für Gesamtzufriedenheit	241
Tabelle 52: Varianzveränderung	243
Tabelle 53: Faktorenladungen	258
Tabelle 54: Übersicht über die Anzahl an Wiederholungen für bestimmte Kinos	278
Tabelle 55: Ladungsstruktur	280
Tabelle 56: Linearer slope – allgemein	281
Tabelle 57: Linearer slope - standortspezifisch	281
Tabelle 58: Frei geschätzter slope	283
Tabelle 59: Lineares vs. frei geschätztes LGC-Modell	284
Tabelle 60: Lineares LGC-Modell – 5 Wellen	284
Tabelle 61: Piecewise LGC-Modell	286
Tabelle 62: Beobachtungsübersicht 12 Wellen	288
Tabelle 63: Wiederholungsmessungen	289
Tabelle 64: Kinospezifisch – fünf Wellen	291
Tabelle 65: Freie Schätzung - standortspezifisch	291

Tabelle 66: Freie Schätzung - Berechnung	292
Tabelle 67: Differenzberechnung	293
Tabelle 68: Konfidenzintervallberechnung – linear und nicht-linear	294
Tabelle 69: Linear vs. nicht-linear mit bzw. ohne indicatorspezifisch zeitübergreifende diachrone Korrelationen	297
Tabelle 70: Fit - Linear vs. nicht-linear mit bzw. ohne indicatorspezifisch zeitübergreifende diachrone Korrelationen	297
Tabelle 71: Berücksichtigung der Zeitabstände	299
Tabelle 72: Berücksichtigung der Zeitabstände und Standorte	299
Tabelle 73: Modellfitevaluierung der Drei-Klassenlösung	317
Tabelle 74: Latente Klassenzugehörigkeitswahrscheinlichkeiten	317
Tabelle 75: Latente Klassenvarianzen	318
Tabelle 76: Klassifikation latenter Klassenmodelle für longitudinale Analysen (Vermunt et al., 2008)	332
Tabelle 77: Klassengrößen - nonstationary manifest Markov Modell	336
Tabelle 78: Transition probabilities	337
Tabelle 79: Stationary manifest vs. stationary latent Markov Modell	338
Tabelle 80: Drei Klassen LCA	339
Tabelle 81: Hidden Markov	341
Tabelle 82: Klassengrößen - Hidden Markov Modell	341
Tabelle 83: Transition matrix (Nylund, 2007)	346
Tabelle 84: Modellüberblick (Muthén, 2006)	349
Tabelle 85: LCA Modellvergleich	350
Tabelle 86: Clusterprofiles	352
Tabelle 87: Parameterspezifikationen im LTA Modell (Muthén und Muthén, 2009)	355
Tabelle 88: Spezifikationsvergleich	356
Tabelle 89: Modellvergleich	357
Tabelle 90: Sechs Zeitpunkte - drei Klassen - stationär - invariantes Messmodell	358
Tabelle 91: Respondentenverteilung über die Klassen	359
Tabelle 92: Sechs Zeitpunkte - zwei Klassen - nicht-stationär - invariantes Messmodell	360
Tabelle 93: Respondentenverteilung über die Klassen	361
Tabelle 94: Longitudinal Guttman Simplex Antwortverhalten (Collins et al., 1998)	369
Tabelle 95: Longitudinal Guttman Simplex (Nylund, 2007)	371
Tabelle 96: Stationäre Übergangswahrscheinlichkeitsmatrix	386
Tabelle 97: Absolute Übergangswahrscheinlichkeiten	386
Tabelle 98: Endgültige Klassenzahlen	387

Tabelle 99: Antwortpatternhäufigkeiten und –anteile des mover-stayer Modells	395
Tabelle 100: Absolute Häufigkeiten der latenten Zustände im mover-stayer Modell	396

8. Formelverzeichnis

Formel 1: Customer Satisfaction-Coefficient (Matzler et al., 1996)	36
Formel 2: Diminishing sensitivity (Mittal et al., 1998)	39
Formel 3: Regressionsmodell (Pollack, 2008)	75
Formel 4: Customer Satisfaction/Dissatisfaction (Bolton und Drew, 1991)	80
Formel 5: Attitude 1 (Bolton und Drew, 1991)	80
Formel 6: Attitude 2 (Bolton und Drew, 1991)	80
Formel 7: Attitude 3 (Bolton und Drew, 1991)	81
Formel 8: Qualität (Bolton und Drew, 1991)	81
Formel 9: Aktuelle Performance (Bolton und Drew, 1991)	81
Formel 10: Statische Empfindung (Bolton und Drew, 1991)	81
Formel 11: Veränderung (Bolton und Drew, 1991)	81
Formel 12: Veränderungsmessung in der Zufriedenheit	83
Formel 13: Paarvergleichskombinationsbewertungen	112
Formel 14: Informationskriterien	128
Formel 15: Modellvergleich	128
Formel 16: AIC und BIC	189
Formel 17: Intra-Class-Korrelation	217
Formel 18: Shrunken Residual	219
Formel 19: Messmodell (Urban, 2004)	265
Formel 20: Strukturmodell (Urban, 2004)	265
Formel 21: Latente Variablen (Preacher et al., 2008)	266
Formel 22: Ein-Item-Modell (Urban, 2004)	266
Formel 23: Konditionales LGC-Modell	272
Formel 24: Bayes-Faktor-Annäherung (zwei Modelle)	303
Formel 25: Bayes Faktor Annäherung (mehr als zwei Modelle)	303
Formel 26: Transition Matrix Felder	347
Formel 27: Markov-Struktur	354

9. Modellverzeichnis

Modell 1: Interceptvariation - Gesamtzufriedenheit	193
Modell 2: Teilzufriedenheiten und Gesamtzufriedenheit	194
Modell 3: Technik/Bild/Ton-Zufriedenheit bezogen auf Gesamtzufriedenheit	195
Modell 4: Random Slope-Modell – Technik/Bild/Ton	197
Modell 5: Nicht-lineares Random Slope-Modell – Technik/Bild/Ton	200
Modell 6: Kundenkartenbesitz und Kinobesuchsverhalten	204
Modell 7: Einfaches Modell	205
Modell 8: Interaktion mit Kinobesuchsverhalten und Kundenkartenbesitz	206
Modell 9: Interaktion mit Kundenkartenbesitz	206
Modell 10: Level-1-Abweichungen	209
Modell 11: Kundenkarteneinfluss auf Besucherebene - Technik/Bild/Ton	210
Modell 12: Kundenkarteneinfluss auf Besucherebene - Technik/Bild/Ton Ergebnis	210
Modell 13: Modell 12 mit Slopevariation	211
Modell 14: Kundenkartenbesitzeinfluss mit Slopevariation - Ergebnisse	212
Modell 15: Kundenkartenvariation	213
Modell 16: Mittelwertschätzung der Standortvariation	215
Modell 17: Standortvariation der Mittelwerte	217
Modell 18: Random Effects-Modell	217
Modell 19: Random Effects-Modell - Ergebnisse	218
Modell 20: Modellvergleich	218
Modell 21: Random Effects-Modell mit drei Ebenen	218
Modell 22: Technik/Bild/Ton-Zufriedenheit und Gesamtzufriedenheit	221
Modell 23: Technik/Bild/Ton-Zufriedenheit und Gesamtzufriedenheit mit Standortabweichungen	221
Modell 24: Einfluss der Cineplexxeigenschaft	223
Modell 25: Ein Level	223
Modell 26: Technik/Bild/Ton-Zufriedenheit zufällig auf Level zwei	224
Modell 27: Level-1- und -2-Einflüsse	229
Modell 28: Interaktion Technik/Bild/Ton-Zufriedenheit und Cineplexx- Eigenschaft	230
Modell 29: Interaktion Besucherverhalten und Technik/Bild/Ton- Zufriedenheit	230
Modell 30: Interaktion Kundenkarte und Technik/Bild/Ton-Zufriedenheit	231
Modell 31: Interaktion Cineplexx-Eigenschaft und Technik/Bild/Ton- Zufriedenheit	231
Modell 32: Entfernung der Kovarianz	233

Modell 33: Random Slope und Intercept Modell	233
Modell 34: Technik/Bild/Ton-Zufriedenheit zur Erklärung der Standortvariabilität	234
Modell 35: Zufälliger Technik/Bild/Ton-Parameter auf Probandenebene	235
Modell 36: Probandenlevelvarianz als Funktion der Besuchereigenschaft und Technik/Bild/Ton-Zufriedenheit	236
Modell 37: Probandenlevelvarianz als Funktion der Cineplexx-Eigenschaft und Technik/Bild/Ton-Zufriedenheit	237
Modell 38: Quadratische und kubische Terme	239
Modell 39: Ordered Multinomial Modell	240
Modell 40: Ordered Multinomial Modell - Schätzwerte	240
Modell 41: Ordered Multinomial Modell auf Level-3-Ebene	241
Modell 42: Einführung eines single variance Terms	242
Modell 43: Grenzen der ordinalen Modellmöglichkeiten	242
Modell 44: Einfaches longitudinales Modell	243
Modell 45: Linear Growth Curve Model	244
Modell 46: Probandenvariation der Wachstumsveränderung	244
Modell 47: Level-1-Variation	246
Modell 48: Nicht-lineares Modell ohne quadratischer Level-2-Variation	246
Modell 49: Nicht-lineares Model mit quadratischer Level-2-Variation	247
Modell 50: Nicht-lineares Modell mit Level-1-Variation	247

10. Literaturverzeichnis

Abu Ghannam, B., Weismayer, C., Wurzer, M., 2009. Zufriedenheit mit Kinos: Anwendung von Paarvergleichs-Pattern-Modellen bei wiederholten Ratings. In: Hatzinger, R., Dittrich, R., Salzberger, T., 2009. Präferenzanalyse mit R: Anwendungen aus Marketing, Behavioural Finance und Human Resource Management, 171-191.

Acock, A., Li, F., 1996. Latent growth curve analysis: a gentle introduction. <http://oregonstate.edu/dept/hdfs/papers/lgcgeneral.pdf>

Acock, A., 2005. Working with missing values. *Journal of Marriage and Family* 67 (4), 1012-1028.

Allan, J., Schattner, P., Stocks, N., Ramsay, E., 2009. Does patient satisfaction of general practice change over a decade? *Family Practice* 10 (13), 1-10.

Anderson, E. W., Mittal, V., 2000. Strengthening the satisfaction-profit chain. *Journal of Service Research* 3 (2), 107-120.

Andrich, D., 1978. A rating formulation for ordered response categories. *Psychometrika* 43 (4), 561-573.

Arnold, M. J., Reynolds, K. E., Ponder, N., Lueg, J. E., 2005. Customer delight in a retail context: investigation delightful and terrible shopping experiences. *Journal of Business Research* 58, 1132-1145.

Audrain-Pontevia, A.F., 2006. Kohonen self-organizing maps: A neural approach for studying the links between attributes and overall satisfaction in a services context. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour* 19, 128-137.

Bernhardt, K. L., Donthu, N., Kennett, P. A., 2000. A longitudinal analysis of satisfaction and profitability. *Journal of Business Research* 47, 161-171.

Bock, R. D., 1972. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37 (1), 29-51.

Bolton, R. N., Drew, J. H., 1991. A longitudinal analysis of the impact of service changes on customer attitudes. *Journal of Marketing* 55 (1), 1-9.

Bühner, M., 2006. Probabilistische Testtheorie. In: Einführung in die Test- und Fragebogenkonstruktion 2., aktualisierte Auflage, Pearson, 300-385.

Buhl, H. U., Kundisch, D., Schackmann, N., Renz, A., 2005. Spezifizierung des Kano-Modells zur Messung von Kundenzufriedenheit. In: Obwerweis, A., Winhardt, C., Gimpel, H., Koschmider, A., Pankratius, V., Schnizler, B., Hrsg., Wirtschaftsinformatik 2007 – eOrganisation: Service-, Prozess-, Market-Engineering, Karlsruhe, Universitätsverlag, 879-912.

Campo, S., Yagüe, M. J., 2009. Exploring non-linear effects of determinants on tourist's satisfaction. *International Journal of Culture, Tourism and Hospitality Research* 3 (2), 127-138.

Celeux, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification* 13, 195-212.

Chen, T.-L., Lee, Y.-H., Hua, C., 2006. Kano two-dimensional quality model and importance-performance analysis in the student's dormitory service quality evaluation in Taiwan. *Journal of American Academy of Business, Cambridge* 9 (2), 324-330.

Chung, H., Park, Y., Lanza, S. T., 2005. Latent transition analysis with covariates: pubertal timing and substance use behaviours in adolescent females. *Statistics in Medicine* 24, 2895-2910.

Collins, L. M., Cliff, N., Dent, C. W., 1988. The longitudinal Guttman simplex: a new methodology for measurement of dynamic constructs in longitudinal panel studies. *Applied Psychological Measurement* 12 (3), 217-230.

Collins, L. M., Cliff, N., 1990. Quantitative methods in psychology: using the longitudinal Guttman simplex as a basis for measuring growth. *Psychological Bulletin* 108 (1), 128-134.

Collins, L. M., Wugalter, S. E., 1992. Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research* 27 (1), 131-157.

Collins, L. M., Graham, J. W., Long, J. D., Hansen, W. B., 1994. Crossvalidation of latent class models of early substance use onset. *Multivariate Behavioral Research* 29 (2), 165-183.

Collins, L. M., Hyatt, S. L., Graham, J. W., 2000. Latent transition analysis as a way of testing models of stage-sequential change in longitudinal data. *Modeling Longitudinal and Multilevel Data: Practical issues, applied approaches, and specific examples*, Lawrence Erlbaum Associates Inc. (Reprint: 2008, Psychology Press), 147-161.

Compton, D., Fuchs, D., Fuchs, L. S., Elleman, A. M., Gilbert, J. K., 2008. Tracking children who fly below the radar: latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences* 18, 329-337.

Coolil, B., Keiningham, T. L. Aksoy, L., Hsu, M., 2007. A longitudinal analysis of customer satisfaction and share of wallet: investigating the moderating effect of customer characteristics. *Journal of Marketing* 71, 67-83.

Crompton, J. L., Duray, N. A., 1985, pre-1986. An investigation of the relative efficacy of four alternative approaches to Importance-Performance Analysis. *Journal of the Academy of Marketing Science* 13 (4), 69-80.

Delucchi, K. L., Matzger, H., Weisner, C., 2004. Dependent and problem drinking over 5 years: a latent class growth analysis. *Drug and Alcohol Dependence* 74, 235-244.

Dittrich, R., Francis, B., Katzenbeisser, W., 2008. Temporal dependence in longitudinal paired comparisons. Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Research Report Series, Report 67, 1-16.

Dittrich, R., Hatzinger, R., 2008. Fitting loglinear Bradley-Terry models (LLBT) for paired comparisons using the R package *prefmod*. Tutorial Part 1, 1-17.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Okut, H., Li, F., 2002. Growth mixture modeling of adolescent alcohol use data. <http://oregonstate.edu/~acock/growth-curves/mixture1-30-02.pdf>

Edwards, D., 2000. *Introduction into graphical modelling*, Springer Verlag, New York, Second Edition.

Embretson, S. E., Reise, S. P., 2000. *Item response theory for psychologists*, Lawrence Erlbaum Associates.

Enders, C. K., Bandalos, D. L., 2001. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal* 8 (3), 430-457.

Finn, A., 2005. Reassessing the foundations of customer delight. *Journal of Service Research* 8 (2), 103-116.

Fischer, G. H., 1989. An IRT-based model for dichotomous longitudinal data. *Psychometrika* 54 (4), 599-624.

Fuchs, M., Weiermair, K., 2004. Destination benchmarking: an indicator-system's potential for exploring guest satisfaction. *Journal of Travel Research* 42, 212-225.

Füller, J., Matzler, K., Faullant, R., 2006. Asymmetric effects in customer satisfaction. *Annals of Tourism Research* 33 (4), 1159-1163.

Fullerton, G., Taylor, S., 2002. Mediating, interactive, and non-linear effects in service quality and satisfaction with services research. *Canadian Journal of Administrative Sciences* 19 (2), 124-136.

Ganglmair, A., Lawson, R., 2003. Measuring affective response to consumption using Rasch modelling. *Journal of Consumer Satisfaction* 16, 198-210.

Ganglmair-Wooliscroft, A., 2007. A comparison of affective response to consumption in two contexts. *Der Markt* 180, 33-46.

Glück, J., Spiel, Ch., 1997. Item-Response-Modelle für Meßwiederholungsdesigns: Anwendung und Grenzen verschiedener Ansätze. *Methods of Psychological Research Online* 2 (1), 1-19.

Hatzinger, R., Mazanec, J. A., 2007. Measuring the part worth of the mode of transport in a trip package: An extended Bradley-Terry model for paired-comparison conjoint data. *Journal of Business Research* 60, 1290-1302.

Hatzinger, R., Rusch, T., 2009. IRT models with relaxed assumptions in eRm: a manual-like instruction. *Psychology Science Quarterly* 51 (1), 87-120.

Johnston, R., 1995. The determinants of service quality: satisfiers and dissatisfiers. *International Journal of Service Industry Management* 6 (5), 53-71.

Jung, T., Wickrama, K. A. S., 2008. An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass* 2 (1), 302-317.

Kano, N., Seraku, N., Takahashi, F., Tsuji, S., 1984. Attractive quality and must-be quality. Report: Miryoku-teki Hinshitu to Atarimue Hinshitu, Hinshitsu. In: *Quality, The Journal of the Japanese Society for Quality Control* 14 (2), 39-48.

Kaplan, D., 2008. An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology* 44 (2), 457-467.

Klein, A. G., Muthen, B. O., 2007. Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research* 42 (4), 647-673.

Kueh, K., 2006. Service satisfiers and dissatisfiers among Malaysian consumers. *Australasian Marketing Journal* 14 (1), 77-90.

Ladhari, R., 2004. The movie experience: A revised approach to determinants of satisfaction. *Journal of Business Research* 60, 454-462.

Langeheine, R., 1988. Manifest and latent Markov chain models for categorical panel data. *Journal of Educational Statistics* 13 (4), 299-312.

Langeheine, R., Van de Pol, F., 2002. Latent Markov chains. In: *Applied Latent Class Analysis*, Chapter 11. Cambridge University Press, 304-341.

Lee, Y.-H., Chen, T.-L., Hua, C., 2006. A Kano two-dimensional quality model in Taiwan's hot spring hotels service quality evaluations. *The Journal of American Academy of Business*, Cambridge 8 (2), 301-305.

de Leeuw, J., Mair, P., (2008). Gifi methods for optimal scaling in R: the package *homals*. Or: Homogeneity analysis in R: the package *homals*. *Journal of Statistical Software* (forthcoming): <http://www.biometrics.mtu.edu/CRAN/web/packages/homals/vignettes/homalsR.pdf>, 1-19.

Li, F., Duncan, T. E., Duncan, S. C., Acock, A., 2001. Latent growth modelling of longitudinal data: a finite growth mixture modelling approach. *Structural Equation Modeling* 8 (4), 493-530.

Luke, D. A., 2004. Multilevel modeling. In: *Series: Quantitative Applications in the Social Sciences*, 1-79.

Mair, P., Hatzinger, R., 2007. Extended Rasch modeling: the eRm package for the application of IRT models in R. *Journal of Statistical Software* 20 (9), 1-20.

Mair, P., Hatzinger, R., 2007. CML based estimation of extended Rasch models with the eRm package in R. *Psychology Science* 49, 26-43.

Mair, R., de Leeuw, J., 2008. Rank and set restrictions for homogeneity analysis in R: the "homals" package. Department of Statistics Papers, University of California, Los Angeles, In: Paper 2008100904: <http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1324&context=uclastat>, 1-10.

Martilla, J. A., James, J. C., 1977, pre-1986. Importance-Performance Analysis. *Journal of Marketing* 41, 77-79.

Maslow, A. H., 1943. A theory of human motivation. *Psychological Review* 50, 370-396.

Masters, G. N., 1982. A Rasch model for partial credit scoring. *Psychometrika* 47 (2), 149-174.

Matzler, K., Bailom, F., Hinterhuber, H.H., Renzl, B., Pichler, J., 2004. The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance-performance analysis. *Industrial Marketing Management* 33 (4), 271-277.

Matzler, K., Hinterhuber, H. H., Bailom, F., Sauerwein, E., 1996. How to delight your customers. *Journal of Product & Brand Management* 5 (2), 6-18.

Matzler, K., Hinterhuber, H. H., 1998. How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation* 18 (1), 25-38.

Matzler, K., Sauerwein, E., 2002. The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis. *International Journal of Service Industry Management* 13 (3/3), 314-332.

Mazanec, J. A., 1979. Probabilistische Meßverfahren in der Marketingforschung: Ein empirischer Anwendungsversuch zur Planung absatzpolitischer Strategien des Imagetransfers. *Marketing – Zeitschrift für Forschung und Praxis* 3, 174-186.

Mazanec, J. A., 2006. Exploring tourist satisfaction with non-linear structural equation modelling and inferred causation analysis. *Journal of Travel & Tourism Marketing* 21 (4), 73-90.

Mazanec, J. A., 2007. “Zauberlehrlings BeSEM” – oder “Was Anwender über Ge- und Missbrauch des Structural Equation Modeling in der betriebswirtschaftlichen Forschung wissen sollten”. *Werbeforschung & Praxis* 1, 25-30.

McDonough, P., Worts, D., Sacker, A., 2009. Socioeconomic inequalities in health dynamics: a comparison of Britain and the United States. *Social Science & Medicine* (70), 251-260.

Meiser, T., Ohrt, B., 1996. Modeling structure and chance in transitions: mixed latent partial markov-chain models. *Journal of Educational and Behavioral Statistics* 21 (2), 91-109.

Michailidis, G., de Leeuw, J., 1998. The Gifi system of descriptive multivariate analysis. *Statistical Science* 13 (4), 307-336.

Mittal, V., Katrichis, J. M., Kumar, P., 2001. Attribute performance and customer satisfaction over time: evidence from two field studies. *The Journal of Services Marketing* 15 (4/5), 343-356.

Mittal, V., Kumar, P., Tsiros, M., 1999. Attribute-level performance, satisfaction, and behavioural intentions over time: a consumption-system approach. *Journal of Marketing* 63, 88-101.

Mittal, V., Ross, W. T., Baldasare, Jr. & P. M., 1998. The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing* 62 (1), 33-47.

Muthén, B., Kaplan, D., 1987. On structural equation modeling with data that are not missing completely at random. *Psychometrika* 52 (3), 431-462.

Muthén, B., Shedden, K., 1999. Finite mixture modelling with mixture outcomes using the EM algorithm. *Biometrics* 55, 463-469.

Muthén, B., Muthén, L. K., 2000. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research* 24 (6), 882-891.

Muthén, B., 2001. Second-generation structural equation modelling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modelling. In: Collins, L. M., Sayer, A.: *New Methods for the Analysis of Change*, APA, Washington, D. C., 291-322.

Muthén, B. O., 2006. Latent variable hybrids: overview of old and new models. Presentation at the University of Maryland CILVR conference "Mixture models in latent variable research", May 18-19.

Muthén, B.O., 2007. Analysis of preventive intervention data using mixture modeling in Mplus. Society for Prevention Research pre-conference workshop, Washington DC, May 29.

Muthén, L. K., Muthén, B. O., 2007. *Mplus user's guide*. Fifth edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., Muthén, B., 2009. Categorical latent variable modeling using Mplus: longitudinal data. *Mplus short courses*, Topic 6.

Nagin, D. S., 1999. Analyzing developmental trajectories: a semiparametric group-based approach. *Psychological Methods* 4 (2), 139-157.

Nylund, K. L., Muthén, B., Nishina, A., Bellmore, A., Graham, S., 2006 (in Press, Manuscript submitted November 30, 2006). Stability and instability of peer victimization during middle school: using latent transition analysis with covariates, distal outcomes, and modelling extensions. In Press, Manuscript submitted November 30, 2006 (http://www.statmodel.com/download/LTA_DP_FINAL.pdf).

Nylund, K. L., 2007. Latent transition analysis: modelling extensions and an application to peer victimization. Dissertation, University of California.

O'Neill, M., 2003. The influence of time on student perceptions of service quality: the need for longitudinal measures. *Journal of Educational Administration* 41 (3), 310-324.

Ostini, R., Nering, M. L., 2006. Polytomous item response theory models. In: *Series: Quantitative Applications in the Social Sciences*, 1-107.

Paulssen, M., Sommerfeld, A., 2006. Modeling the nonlinear relationship between satisfaction and loyalty with structural equation models. In: Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberger, A., Gaul, W.. *Studies in classification, data analysis, and knowledge organization: From Data and Information Analysis to Knowledge Engineering*, 574-581.

Poinstingl, H., Mair, P., Hatzinger, R., 2007. Manual zum Softwarepackage eRm (extended Rasch modelling): Anwendung des Rasch-Modells (1-PL Modell), http://erm.r-forge.r-project.org/eRm_manual.pdf, 1-131.

Pollack, B. L., 2008. The nature of the service quality and satisfaction relationship: empirical evidence for the existence of satisfiers and dissatisfiers. *Managing Service Quality* 18 (6), 537-558.

Poulsen, C. S., 1990. Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing* 7, 5-19.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., Briggs, N. E., 2008. Latent growth curve modeling. In: *Series: Quantitative Applications in the Social Sciences*, 1-96.

Rasbash, J., Steele, F., Browne, W., Prosser, B., 2005. A user's guide to MLwiN, Version 2.0. Centre for Multilevel Modelling, University of Bristol, 1-256.

Reddy, R., Rhodes, J. E., Mulhall, P., 2003. The influence of teacher support on student adjustment in the middle school years: a latent growth curve study. *Development and Psychopathology* 15, 119-138.

Rubin, D., 1976. Inference and missing data. *Biometrika* 63 (3), 581-592.

Sampson, S. E., Showalter, M. J., 1999. The Performance-Importance Response Function: Observations and Implications. *The Service Industries Journal* 19 (3), 1-25.

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., Kellam, S., 2003. Modeling growth in boys' aggressive behaviour across elementary school: links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology* 39 (6), 1020- 1035.

Steele, F., 2008. Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A* 171 (Part 1), 5-19.

Stoel, R. D., van Den Wittenboer, G., Hox, J., 2003. Analyzing longitudinal data using multilevel regression and latent growth curve analysis. *Metodología de las Ciencias del Comportamiento* 5, 21-42.

Torres, E. N., Kline, S., 2006. From satisfaction to delight: a model for the hotel industry. *International Journal of Contemporary Hospitality Management* 18 (4), 290-301.

Urban, D., 2004. *Neue Methoden der Längsschnittanalyse: Zur Anwendung von latenten Wachstumskurvenmodellen in Einstellungs und Sozialisationsforschung*. LIT Verlag Münster, 1-146.

Van de Pol, F., Langeheine, R., 1990. Mixed Markov latent class models. *Sociological Methodology* 20, 213-247.

Van de Pol, F., Mannan, H., 2002. Questions of a novice in latent markov modelling. *Methods of Psychological Research Online*, Special Issue, 1-18.

Van Ittersum, K., Pennings, J. M. E., Wansink, B., van Trip, H. C. M., 2007. The validity of attribute-importance measurement: A review. *Journal of Business Research* 60, 1177-1190.

Vargo, S. L., Nagao, K., He, Y., Morgan, F. W., 2007. Satisfiers, dissatisfiers, criticals, and neutrals: a review of their relative effects on customer (dis)satisfaction. *Academy of Marketing Science Review* 11 (2), 1-19.

Vermunt, J. K., Langeheine, R., Bockenholt, U., 1999. Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics* 24 (2), 179-207.

Vermunt, J. K., 2004. Mover-stayer models. In: Encyclopedia of Research Methods for the Social Sciences. Edited by Liao, T. F., New Bury Park: Sage Publications.

Vermunt, J. K., 2007. Growth models for categorical response variables: standard, latent-class, and hybrid approaches. In: van Montfort, K., Oud, J., Satorra, A.: Longitudinal models in the behavioural and related sciences, 2007. Publisher: Lawrence Erlbaum Associates, Mahwah, New Jersey, 139-158.

Vermunt, J. K., Tran, B., Magidson, J., 2008. Latent class models in longitudinal research. In: S. Menard (ed.), Handbook of Longitudinal Research: Design, Measurement, and Analysis, Burlington, MA: Elsevier, 375-385.

Weismayer, Ch., 2010. Longitudinal satisfaction measurement using latent growth curve models and extensions. Journal of Retailing and Consumer Services 17 (4), 321-331.

Forschungsergebnisse der Wirtschaftsuniversität Wien

Herausgeber: Wirtschaftsuniversität Wien –
vertreten durch a.o. Univ. Prof. Dr. Barbara Sporn

INFORMATION UND KONTAKT:

WU (Wirtschaftsuniversität Wien)
Department of Finance, Accounting and Statistics
Institute for Finance, Banking and Insurance
Heiligenstädter Straße 46-48, 1190 Wien
Tel.: 0043-1-313 36/4556
Fax: 0043-1-313 36/904556
valentine.wendling@wu.ac.at
www.wu.ac.at/finance

- Band 1 Stefan Felder: Frequenzallokation in der Telekommunikation. Ökonomische Analyse der Vergabe von Frequenzen unter besonderer Berücksichtigung der UMTS-Auktionen. 2004.
- Band 2 Thomas Haller: Marketing im liberalisierten Strommarkt. Kommunikation und Produktplanung im Privatkundenmarkt. 2005.
- Band 3 Alexander Stremitzer: Agency Theory: Methodology, Analysis. A Structured Approach to Writing Contracts. 2005.
- Band 4 Günther Sedlacek: Analyse der Studiendauer und des Studienabbruch-Risikos. Unter Verwendung der statistischen Methoden der Ereignisanalyse. 2004.
- Band 5 Monika Knassmüller: Unternehmensleitbilder im Vergleich. Sinn- und Bedeutungsrahmen deutschsprachiger Unternehmensleitbilder – Versuch einer empirischen (Re-)Konstruktion. 2005.
- Band 6 Matthias Fink: Erfolgsfaktor Selbstverpflichtung bei vertrauensbasierten Kooperationen. Mit einem empirischen Befund. 2005.
- Band 7 Michael Gerhard Kraft: Ökonomie zwischen Wissenschaft und Ethik. Eine dogmenhistorische Untersuchung von Léon M.E. Walras bis Milton Friedman. 2005.
- Band 8 Ingrid Zechmeister: Mental Health Care Financing in the Process of Change. Challenges and Approaches for Austria. 2005.
- Band 9 Sarah Meisenberger: Strukturierte Organisationen und Wissen. 2005.
- Band 10 Anne-Katrin Neyrer: Multinational teams in the European Commission and the European Parliament. 2005.
- Band 11 Birgit Trukeschitz: Im Dienst Sozialer Dienste. Ökonomische Analyse der Beschäftigung in sozialen Dienstleistungseinrichtungen des Nonprofit Sektors. 2006
- Band 12 Marcus Kölling: Interkulturelles Wissensmanagement. Deutschland Ost und West. 2006.
- Band 13 Ulrich Berger: The Economics of Two-way Interconnection. 2006.
- Band 14 Susanne Guth: Interoperability of DRM Systems. Exchanging and Processing XML-based Rights Expressions. 2006.
- Band 15 Bernhard Klement: Ökonomische Kriterien und Anreizmechanismen für eine effiziente Förderung von industrieller Forschung und Innovation. Mit einer empirischen Quantifizierung der Hebeleffekte von F&E-Förderinstrumenten in Österreich. 2006.

- Band 16 Markus Imgrund: Wege aus der Insolvenz. Eine Analyse der Fortführung und Sanierung insolventer Klein- und Mittelbetriebe unter besonderer Berücksichtigung des Konfigurationsansatzes. 2007.
- Band 17 Nicolas Knotzer: Product Recommendations in E-Commerce Retailing Applications. 2008.
- Band 18 Astrid Dickinger: Perceived Quality of Mobile Services. A Segment-Specific Analysis. 2007.
- Band 19 Nadine Wiedermann-Ondrej: Hybride Finanzierungsinstrumente in der nationalen und internationalen Besteuerung der USA. 2008.
- Band 20 Helmut Sorger: Entscheidungsorientiertes Risikomanagement in der Industrieunternehmung. 2008.
- Band 21 Martin Rietsch: Messung und Analyse des ökonomischen Wechselkursrisikos aus Unternehmenssicht: Ein stochastischer Simulationsansatz. 2008.
- Band 22 Hans Christian Mantler: Makroökonomische Effizienz des Finanzsektors. Herleitung eines theoretischen Modells und Schätzung der Wachstumsimplikationen für die Marktwirtschaften und Transformationsökonomien Europas. 2008.
- Band 23 Youri Tacoun: La théorie de la valeur de Christian von Ehrenfels. 2008.
- Band 24 Monika Koller: Longitudinale Betrachtung der Kognitiven Dissonanz. Eine Tagebuchstudie zur Reiseentscheidung. 2008.
- Band 25 Marcus Scheiblecker: The Austrian Business Cycle in the European Context. 2008.
- Band 26 Aida Numic: Multinational Teams in European and American Companies. 2008.
- Band 27 Ulrike Bauernfeind: User Satisfaction with Personalised Internet Applications. 2008.
- Band 28 Reinhold Schodl: Systematische Analyse und Bewertung komplexer Supply Chain Prozesse bei dynamischer Festlegung des Auftragsentkopplungspunkts. 2008.
- Band 29 Bianca Gusenbauer: Öffentlich-private Finanzierung von Infrastruktur in Entwicklungsländern und deren Beitrag zur Armutsreduktion. Fallstudien in Vietnam und auf den Philippinen. 2009.
- Band 30 Elisabeth Salomon: Hybrides Management in sino-österreichischen Joint Ventures in China aus österreichischer Perspektive. 2009.
- Band 31 Katharina Mader: Gender Budgeting: Ein emanzipatorisches, finanzpolitisches und demokratiepolitisches Instrument. 2009.
- Band 32 Michael Weber: Die Generierung von Empfehlungen für zwischenbetriebliche Transaktionen als gesamtwirtschaftliche Infrastrukturleistung. 2010.
- Band 33 Lisa Gimpl-Heersink: Joint Pricing and Inventory Control under Reference Price Effects. 2009.
- Band 34 Erscheint nicht.
- Band 35 Dagmar Kiefer: Multicultural Work in Five United Nations Organisations. An Austrian Perspective. 2009.
- Band 36 Gottfried Gruber: Multichannel Management. A Normative Model Towards Optimality. 2009.
- Band 37 Rainer Quante: Management of Stochastic Demand in Make-to-Stock Manufacturing. 2009.
- Band 38 Franz F. Eiffe: Auf den Spuren von Amartya Sen. Zur theoriegeschichtlichen Genese des Capability-Ansatzes und seinem Beitrag zur Armutsanalyse in der EU. 2010.

- Band 39 Astrid Haider: Die Lohnhöhe und Lohnstreuung im Nonprofit-Sektor. Eine quantitative Analyse anhand österreichischer Arbeitnehmer-Arbeitgeber-Daten. 2010.
- Band 40 Maureen Lenhart: Pflegekräftemigration nach Österreich. Eine empirische Analyse. 2010.
- Band 41 Oliver Schwank: Linkages in South African Economic Development. Industrialisation without Diversification? 2010.
- Band 42 Judith Kast-Aigner: A Corpus-Based Analysis of the Terminology of the European Union's Development Cooperation Policy, with the African, Caribbean and Pacific Group of States. 2010.
- Band 43 Emel Arıkan: Single Period Inventory Control and Pricing. An Empirical and Analytical Study of a Generalized Model. 2011.
- Band 44 Gerhard Wohlgenannt: Learning Ontology Relations by Combining Corpus-Based Techniques and Reasoning on Data from Semantic Web Sources. 2011.
- Band 45 Thomas Peschta: Der Einfluss von Kundenzufriedenheit auf die Kundenloyalität und die Wirkung der Wettbewerbsintensität am Beispiel der Gemeinschaftsverpflegungsgastronomie. 2011.
- Band 46 Friederike Hehle: Die Anwendung des Convenience-Konzepts auf den Betriebstyp Vending. 2011.
- Band 47 Thomas Herzog: Strategisches Management von Koopetition. Eine empirisch begründete Theorie im industriellen Kontext der zivilen Luftfahrt. 2011.
- Band 48 Christian Weismayer: Statische und longitudinale Zufriedenheitsmessung. 2011.
- Band 49 Johannes Fichtinger: The Single-Period Inventory Model with Spectral Risk Measures. 2011.
- Band 50 Isabella R. Hatak: Kompetenz, Vertrauen und Kooperation. Eine experimentelle Studie. 2011.

