



11. Tagung der Fachgruppe Methoden & Evaluation
der Deutschen Gesellschaft für Psychologie,

19.-21. September 2013,
Alpen-Adria-Universität Klagenfurt.

Abstracts

Modelltests revisited

Alexandrowicz, Rainer W.

rainer.alexandrowicz@aau.at

Nachdem das Problem der Identifikation der H1-Verteilung für den Wald-Test zur Überprüfung des Rasch-Modells nun gelöst ist, stellt sich die Frage nach einem geeigneten Effektstärkemaß, um optimale Stichprobenplanung auch in der Praxis umsetzen zu können. Die vorgestellte Studie untersucht verschiedene Herangehensweisen hinsichtlich ihres praktischen Nutzens.

Session 14 – IRT V: Modelltests & Parameterschätzung (Fr, 20.9., 15:40–17:00, HS C)

.....

Using Testlet Response Theory to Examine Local Dependency in C-Tests

Baghaei, Purya, Thomas Eckes

puryabaghaei@gmail.com

The conditional independence assumption made in Rasch and IRT models requires test items to be independent conditional on the latent trait, i.e., there should not be any correlation between the items after conditioning on the latent trait. Items sharing a common stimulus as in testlets may violate this assumption and measure some other factor beyond the intended latent trait. Basically two approaches have been suggested to model and account for the dependency among items nested within a testlet: (a) the scores of the items based on the same testlet are added and the sum scores are treated as categories of a polytomous item that is analyzed with polytomous Rasch or IRT models (item-bundle approach; Rosenbaum, 1988), (b) the local dependency between items of a given testlet is explicitly modeled such as in testlet response theory (TRT; Wainer, Bradlow, & X. Wang, 2007) or related approaches. In this study we examined local dependency in a gap-filling language proficiency test that followed the testlet-based format of C-tests. We used both approaches and compared the resulting parameter estimates with those obtained from a standard IRT model. Testlet effects proved to be very small and, therefore, there was a high degree of correspondence between parameters obtained from the different modeling approaches. Findings, however, also indicated that when local dependence was ignored reliability of the C-test was overestimated. Implications for the analysis of testlets in general, and C-tests in particular are discussed.

Session 6 – IRT I: Testdesign (Do, 19.9., 15:20–17:00, HS C)

.....

Welches Verfahren findet die richtige Anzahl an Dimensionen? Vergleich von EFA und CA

Bollmann, Stella, Heene, Moritz; Bühner, Markus
stella.bollmann@googlemail.com

Die vorliegende Studie untersuchte die Effektivität der exploratorischen Faktorenanalyse (EFA) und der Clusteranalyse (CA) beim Auffinden der Anzahl an Dimensionen. Neben simulierten Daten wurde ein realer, großer Datensatz analysiert, um die Probleme normaler Simulationsstudien zu umgehen. Als Kriterium für die wahre Anzahl an Dimensionen wurde das Populationsmodell dieses Datensatzes herangezogen. Dann zogen wir Stichproben verschiedener Größe aus diesem Datensatz und untersuchten, wie oft die jeweiligen Methoden die gleiche Anzahl an Dimensionen ergaben. Schließlich wurden diese Ergebnisse mit den Ergebnissen aus Simulationsbedingungen verglichen, in denen Nebenladungen, Faktorinterkorrelationen und Stichprobengrößen variiert wurden. Die Ergebnisse zeigen, dass sowohl die Parallelanalyse für Hauptkomponentenanalyse (PA-PAF) als auch der MAP-Test für größere Stichproben ($n=150$) zufriedenstellende Ergebnisse liefern. Ihre Effektivität wurde lediglich von einer CA Methode erreicht, der average linkage mit Korrelation der Korrelation (AL-RCC). Sowohl AIC als auch BIC waren stark abhängig von der Stichprobengröße. Von hohen Nebenladungen war am meisten die Effektivität von PA mit Hauptkomponentenanalyse (PA-PCA) und am wenigsten die von PA-PAF beeinträchtigt. Alle Verfahren zeigten deutlich bessere Ergebnisse in den Simulationen als im realen Datensatz auch wenn in der Simulation Nebenladungen und Faktorinterkorrelationen aus dem realen Datensatz übernommen wurden.

Session 4 – Regression – FA – VA – CA (Do, 19.9., 13:20–15:00, HS C)

.....

Auswirkung schwierigkeitsbestimmender Strukturmerkmale bei der mehrfach restringierten Aufgabenauswahl im Kontext des multidimensionalen adaptiven Testen (MAT)

Born, Sebastian, Frey, Andreas
sebastian.born@uni-jena.de

Die beim MAT eingesetzten Algorithmen zur Itemselektion folgen meist nur statistischen Kriterien (z.B. maximaler Information). Dies kann dazu führen, dass gewünschte nicht-statistische Restriktionen nicht erfüllt werden. Mit dem von Frey, Cheng und Seitz (2010) vorgeschlagenen Multidimensional-Maximum-Priority-Index (MMPI) ist es möglich, zusätzliche nicht-statistische Restriktionen hinsichtlich der Strukturmerkmale simultan zu inhaltlichen Restriktionen zu berücksichtigen. Ungeklärt dabei sind die Auswirkungen schwierigkeitsbestimmender Strukturmerkmale bei Verwendung des MMPI zur mehrfach restringierten Aufgabenauswahl. Im Rahmen einer Simulationsstudie soll dies näher untersucht werden. Der Studie liegt ein dreifaktorielles Design mit den Faktoren Abhängigkeit der Itemschwierigkeit von den Strukturmerkmalen (Abhängigkeit, Unabhängigkeit), Selektionsalgorithmus (MAT, MAT+MMPI) und der Anzahl der Restriktionen zu Grunde. Evaluationskriterien für die Anwendbarkeit des MMPI sind die Messpräzision und das Ausmaß, in dem die Restriktionen eingehalten werden. Es ist zu erwarten, dass die restringierte Auswahl (MMPI) in der Abhängigkeitsbedingung im Vergleich zur Unabhängigkeitsbedingung zu einer Verringerung der Messpräzision in den Randbereichen der Fähigkeitsverteilung führt. Bei unrestringierter Auswahl (MAT) sollten sich dagegen keine Unterschiede in der Messpräzision zeigen. Vor dem Hintergrund der Ergebnisse wird die Anwendbarkeit der erweiterten MMPI-Methode diskutiert.

Session 6 – IRT I: Testdesign (Do, 19.9., 15:20–17:00, HS C)

.....

The Estimation of latent interaction and quadratic effects under the condition of non-normally distributed data

Brandt, Holger, Kelava, Augustin und Klein, Andreas
holger.brandt@uni-tuebingen.de

Methods for the estimation of latent interaction and quadratic effects in SEM are typically based on a normality assumption for all or most of the variables in the model (e.g., for product indicator approaches or LMS). This assumption is often violated in empirical situations. Therefore, these methods may lead to biased estimates and incorrect inferential decisions when data are non-normal. Recently, some methods have been proposed to estimate latent interaction and quadratic effects particularly in situations when data are non-normal. However, these methods have not yet been available in standard statistics software. Here, we present a simulation study that investigates the performance of two method-of-moments approaches – the 2SMM approach (Wall & Amemiya, 2000, 2003) and the MM approach (Mooijjaart & Bentler, 2010) – and a nonlinear structural equation mixture modeling approach (NSEMM approach; Kelava, Nagengast & Brandt, in press) in comparison to the LMS estimator and the extended unconstrained approach under the condition of normal and non-normal data. The results suggest that LMS performs best as long as the data are normally distributed. Under the condition of non-normal data, the 2SMM approach and the NSEMM approach perform best and should be used in these situations. In contrast, the MM approach leads to biased estimates and strongly inflated type one error rates. Limitations of the approaches and future extensions are discussed.

Session 7 – SEM II: Nichtlineare Effekte (Fr, 20.9., 10:00–11:00, HS B)

.....

Modellierung des Antwortformats als genesteten Faktor in einem MIRT-Modell mit latenter Interaktion

Buchholz, Janine, Hartig, Johannes
buchholz@dipf.de

Die Modelle der Mehrdimensionalen Item Response Theorie (MIRT) mit Mehrfachladung unterscheiden sich im Zusammenspiel der beteiligten Dimensionen zur Vorhersage der Lösungswahrscheinlichkeit einzelner Items. Beispielsweise kann eine Interaktion zweier Dimensionen (z.B. Fähigkeiten) eingefügt werden, sodass die Lösungswahrscheinlichkeit erst dann ansteigt, wenn beide Fähigkeiten zunehmend vorliegen. In einer vorangegangenen Simulationsstudie konnte bereits gezeigt werden, dass dieses Interaktionsmodell sogar dann geschätzt werden kann, wenn eine der beiden Dimensionen genestet ist, also keine reinen Indikatoren besitzt. Dieses Modell wurde nun zur Modellierung von schriftlichen Mathematikaufgaben aus PISA 2009 herangezogen, in denen offene und geschlossene Antwortformate vorlagen. Es wird gezeigt, dass zur Lösung von Items mit offenem Antwortformat neben allgemeiner Kompetenz in Mathematik eine zusätzliche Fähigkeit erforderlich ist, die bereits als Mathematisches Kommunizieren beschrieben wurde und im Interaktionsmodell den zweiten, genesteten Faktor darstellt. Somit weist das Interaktionsmodell eine bessere Modellgüte gegenüber einem kompensatorischen Modell auf, in dem Mathematikkompetenz und Mathematisches Kommunizieren additiv zur Lösungswahrscheinlichkeit beitragen.

Session 10 – IRT III: Antwortformate & Itempolung (Fr, 20.9., 11:20–12:40, HS C)

.....

Eine Vollerhebung und methodologische Analyse von Onlinefragebögen auf den fünf größten internationalen Online-Rekrutierungsportalen.

Burger, Christoph, Stieger, Stefan
a0307384@unet.univie.ac.at

Die Anwendung von Online-Erhebungsmethoden gewinnt in der Psychologie zunehmend an Bedeutung und die methodologische Literatur zu diesem Thema wächst kontinuierlich. Bisherige Studien zur Datenqualität zeigten beispielsweise, dass bestimmte Onlinefragebogen-Merkmale wegen ihrer unvorteilhaften Effekte vermieden werden sollten (z.B. forced response), während bestimmte andere Merkmale empfohlen werden, weil sie zu besserer Datenqualität führen (z.B. Fortschrittsanzeigen). Bisher ist jedoch weithin unbekannt, ob diese Empfehlungen in der aktuellen Forschungspraxis auch tatsächlich umgesetzt werden. Die vorliegende Arbeit hatte daher zum Ziel, die momentanen Praktiken bei Onlineerhebungen aus einer Metaebene überblicksartig darzustellen und Empfehlungen für die Anwendung zukünftiger Online-Forschungsprojekte abzuleiten. Die Datengrundlage war eine Vollerhebung aller verfügbaren Onlinefragebögen (N = 720) auf fünf hochfrequentierten internationalen Onlineportalen zur Zeit der Datenerhebung (Frühjahr 2012). Es zeigte sich, dass die Onlinestudien bei der Vermeidung von unvorteilhaften Merkmalen (z.B. vorausgefüllte Antwortoptionen) gut abschnitten, während sie beim empfohlenen Einsatz von vorteilhaften Merkmalen (z.B. Fortschrittsanzeigen) äußerst schlecht abschnitten. Zukünftige Onlinestudien sollten daher vermehrt auf die Implementierung von vorteilhaften Merkmalen in Onlinefragebögen setzen.

Session 16 – IRT VI: Anwendungen / Datenerhebung (Sa, 21.9., 9:30–10:50, HS C)

.....

Modellierung unterschiedlicher Antwortformate in einparametrischen IRT Modellen

Carstensen, Claus, Haberkorn, Kerstin, Pohl, Steffi
claus.carstensen@uni-bamberg.de

Im Nationalen Bildungspanel (NEPS) werden bei der Messung von Kompetenzen im Wesentlichen zwei Antwortformate verwendet, einfache Multiple-Choice (MC) Items und komplexe Multiple-Choice (CMC) Items. Wie in vielen Studien werden die Teilaufgaben komplexer Multiple-Choice Items aggregiert und mittels Partial Credit Modellen (PCM) analysiert, während die Antworten aus MC-Aufgaben mit dem dichotomen Raschmodell (RM) modelliert werden. Im Bildungspanel werden mit dem RM und dem PCM einparametrische Antwortmodelle zur Skalierung eingesetzt, die keine unterschiedlichen Trennschärfen der Aufgaben modellieren. Durch die Mischung der beiden Antwortformate ergeben sich unterschiedliche Höchstpunktzahlen für verschiedene Aufgaben in einem Testinstrument. Unterschiedliche Punktzahlen für die Antworten in einem Instrument implizieren aber unterschiedliche Trennschärfen der hier eingesetzten Antwortformate. Im vorgeschlagenen Beitrag werden verschiedene Regeln zur Vergabe von Punkten vorgestellt und es wird untersucht, ob und wie sich Antworten unterschiedlicher Antwortformate auf eine gemeinsame Skala abbilden lassen, ohne dass für jede Aufgabe ein eigener Trennschärfeparameter wie im 2PLM angenommen werden muss. Eines der zentralen Ergebnisse deutet darauf hin, dass die CMC Items mit einem halben Punkt je richtiger Teilaufgaben auf derselben Skala messen wie die dichotom bewerteten Aufgaben in den untersuchten Testinstrumenten.

Session 10 – IRT III: Antwortformate & Itempolung (Fr, 20.9., 11:20–12:40, HS C)

.....

Ein Modell für Zählraten aus Tests unter Zeitdruck: Kann man Schwierigkeit und Geschwindigkeit trennen?

Doebler, Anna, Doebler, Philipp; Holling, Heinz
anna.doebler@uni-muenster.de

Wenn Aufgaben in einem Test sehr einfach sind, so könnenviele davon in kurzer Zeit bearbeitet werden. So werden z.B. inAufmerksamkeits- und Konzentrationstests, aber auch in Lesetests, oft viele Stimuli mit einem Zeitlimit präsentiert und nur die Gesamtzahl der richtig bearbeiteten Stimuli fließt dann ins Scoring ein. Rasch führte 1960 in seiner bekannten Monographie das Poisson Counts Modell für Zählraten aus solchen Tests ein, welches im Kern log-linear mit Poisson-Variation ist und einen Offset für die bekannte Bearbeitungszeit enthält. Items (im Sinne von Bündeln von Stimuli) haben jeweils nur einen zu schätzenden Parameter, und somit kann das Poisson Counts Modell nicht zwischen Schwierigkeit und Geschwindigkeit von Items trennen. In diesem Beitrag geht es um ein Modell, welches auf einer logistischen Itemcharakteristischen Kurve aufbaut und auch eine Poisson Verteilung für die beobachteten Scores annimmt. Da das Modell aber für jedes Item sowohl einen Schwierigkeits- als auch einen Geschwindigkeitsparameter enthält, kann man damit feiner modellieren. Das Rasch Poisson Counts Modell ist als Sonderfall enthalten: es ergibt sich als Modell mit konstanten Itemschwierigkeiten. Zur Parameterschätzung nutzt man z.B. ein joint maximum likelihood Verfahren. Wir wenden das Modell auf Daten aus einem Test für Bearbeitungs- und Geschwindigkeit an, in welchem sich Schwierigkeit und Geschwindigkeit mit Hilfe des neuen Modells trennen lassen.

Session 16 – IRT VI: Anwendungen / Datenerhebung (Sa, 21.9., 9:30–10:50, HS C)

.....

Idiographische Prozessforschung in der Psychotherapie und der Suizidprävention

Fartacek, Clemens, Plöderl Martin, , Fartacek Reinhold, Schiepek Günter
Guenter.schiepek@ccsys.de

Neben standardisierten Messinstrumenten zur Erfassung von Outcome und Prozess (für tägliche Selbsteinschätzungen hat sich insbesondere der Therapie-Prozessbogen [TPB] bewährt) eröffnen neue Internet-basierte Monitoring-Systeme auch die Möglichkeit, patientenspezifische, individualisierte Fragebögen zu nutzen. Im Synergetischen Navigationssystem (SNS) steht hierfür ein Fragebogeneditor zur Verfügung. Die Items der idiographischen Fragebögen werden auf der Basis qualitativer Interviews (Ressourceninterview) und Strukturentwicklungsverfahren (idiographische Systemmodellierung) gemeinsam mit dem Patienten entwickelt. Die Dynamiken der entsprechenden Therapieprozesse können damit zeitlich wie inhaltlich hochspezifisch erfasst und in den Therapieprozess zurückgemeldet werden. Diskutiert werden die methodischen Fragen der Validität und der Reaktivität solcher individualisierter therapie-katalysierender Methoden. Erste Schritte wurden auf dieser Basis auch in Richtung eines Frühwarnsystems für suizidale Krisen gemacht, wobei in den "real-time", d.h. in Tagestaktung anfallenden Zeitreihendaten spezifische "Precursors" für solche Krisen identifizierbar sind.

Session 2 – Prozessmonitoring & fMRI (Do, 19.9., 9:20–11:00, HS C)

.....

Evaluation eines Studiums mit Mixed Methods

Fenzl, Thomas, Philipp Mayring

Thomas.Fenzl@aau.at

Der im Jahr 2005 an der IFF-Fakultät der Universität Klagenfurt eingeführte Master-Studiengang Sozial- und Humanökologie wurde mit Beginn des WS 2009/10 einer internen, formativ orientierten Evaluation unterzogen. Basierend auf den in den Verbundevaluationen in Deutschland entwickelten Kriterien zur Evaluation von Studium und Lehre wurden in den Vorgesprächen die Einschätzung der Berufschancen der Absolventen/innen, die Perspektiven der Entwicklung der Studierendenzahlen, die Studierendenzufriedenheit sowie eine Absolventen/innen-Befragung ins Zentrum der Evaluation gerückt. Zu diesem Zweck wurden Studiengangs-Kennzahlen analysiert und in Dokumentenanalysen die Einbettung in die internationale Forschungslandschaft geklärt. Des Weiteren wurden Studierenden- und AbsolventInnen-Befragungen durchgeführt und Interviews mit Institutsangehörigen, Lehrbeauftragten und Vertreter/innen (z.B. Personalverantwortliche) der vom Studiengang angepeilten Berufsfelder geführt. Die gesammelten Daten wurden anhand der Bewertungskriterien eingeschätzt und zu einem Bericht zusammengestellt. Ergänzt wurde diese Perspektive durch die Ergebnisse des Selbstevaluationsprozesses, welchem sich das Institut für Soziale Ökologie unterzogen hat. Der Evaluationsbericht und der Selbstbericht bildeten die Grundlage für die Begutachtung des Masterstudiums durch zwei GutachterInnen und dienten dem internen Prozess der Qualitätssicherung sowie zur Rechenschaftslegung nach außen.

Session 13 – Mixed Methods (Fr, 20.9., 15:40–17:00, HS B)

.....

Identifying moderated quadratic effects in structural equation models

Fischer, Kevin Alexander, Schermelleh-Engel, Karin; Moosbrugger, Helfried; Klein, Andreas

kefische@stud.uni-frankfurt.com

For the analysis of latent nonlinear relations researchers are usually investigating moderated linear effects and/or quadratic effects. However, there are relations between variables which cannot be approximated by linear relations. One example is a moderated quadratic effect which is present when the quadratic relation between the predictor and the criterion differs depending on the level of the moderator variable. Latent structural equations (LMS) can be used for the analysis of moderated quadratic effects. However, in order to do so all terms of the same order and below have to be included in the equation. Unfortunately, this is associated with an increased need of sample size to detect the moderated quadratic effect. We conducted a Monte Carlo study in which we varied the size of the moderated quadratic effect and the number of nonlinear terms. The results show that the power of LMS decreases the more nonlinear terms have to be included in the equation. We therefore developed a new screening procedure as an attempt to identify all nonlinear terms of the structural equation, to avoid the inclusion of inconsiderable nonlinear terms in LMS which are not present in the sample. The reduction of nonlinear terms results in an increased power to detect a moderated quadratic effect. The benefits and limitations of the screening procedure will be discussed.

Session 7 – SEM II: Nichtlineare Effekte (Fr, 20.9., 10:00–11:00, HS B)

.....

Modellierung von Lehrevaluationsdaten unter Berücksichtigung der Mehrebenenstruktur

Fondel, Eva, Pfeiffer, Andreas; Weis, Susanne; Lischetzke, Tanja; Gollwitzer, Mario
eva.fondel@gmx.de

Im Rahmen der universitären Qualitätssicherung sind Lehrveranstaltungsevaluationen anhand von Studierendenbefragungen mit standardisierten Fragebögen eine gängige Praxis. In den meisten Analysen zu Messeigenschaften und Gütekriterien der Evaluationsinstrumente werden Korrelationen oder exploratorische Faktorenanalysen, in seltenen Fällen auch konfirmatorische Faktorenanalysen herangezogen. Die notwendige Berücksichtigung der in den Daten innewohnenden Mehrebenenstruktur bleibt dabei oft außen vor. Die vorliegende Studie adressiert diese methodischen Probleme und zeigt eine Modellierung solcher Daten in einem Multilevel confirmatory factor analysis - multitrait-multimethod (ML-CFA-MTMM) model for interchangeable raters (Eid et al., 2008). Durch eine Varianzaufteilung in Traitfaktoren (Lehrendenebene), Methodenfaktoren (Studierendenebene) und Messfehler ermöglicht dieses Modell eine um verschiedene Störeinflüsse bereinigte Abbildung des Zielkonstrukts Lehrqualität. Des Weiteren können differenzierte Aussagen zur Reliabilität, konvergenten Validität und Methodenspezifität getroffen werden. Die Studie basiert auf N= 3194 Studierendenratings in 169 Lehrveranstaltungen der Universität Koblenz-Landau. Evaluationsinstrument ist eine adaptierte Variante des Trierer Inventars zur Lehrevaluation (TRIL; Gollwitzer & Schlotz, 2003). Zusätzlich wurden häufig diskutierte Biasvariablen wie u.a. Interesse der Studierenden und die Sympathie für den Lehrenden erhoben.

Session 15 – Evaluation II (Sa, 21.9., 9:30–10:50, HS B)

.....

Zur Technik des lauten Denkens als Methode der Usability-Evaluation

Forthmann, Boris
boris.forthmann@uni-muenster.de

Um die Usability einer Software einzuschätzen, wird im Rahmen von Usability-Evaluationen (UE) häufig auf die Methode des lauten Denkens (MLD) zurückgegriffen. Bei den in der MLD geäußerten Verbalisierungen wird zwischen Level-1-Verbalisierungen (Verbalisierungen die in verbaler Form im augenblicklichen Aufmerksamkeitsfokus einer Person liegen), Level-2- Verbalisierungen (Verbalisierungen die im augenblicklichen Aufmerksamkeitsfokus einer Person liegen, aber zuvor noch ins verbale Format gebracht werden müssen) und Level-3- Verbalisierungen (Verbalisierungen, die geäußert werden, werden durch eine Verschiebung des Aufmerksamkeitsfokus auf frühere Gedanken aus dem Langzeitgedächtnis beeinflusst) unterschieden. Dabei werden nach theoretischen Annahmen nur Level-1- und Level-2- Verbalisierungen als valides Abbild der ablaufenden kognitiven Prozesse verstanden. Um die Rolle der Art der Verbalisierungen im Rahmen der UE zu klären, wurden $n = 1725$ Verbalisierungen im Rahmen einer UE-Studie kodiert. Mithilfe multi-level log-linearer Modelle und lag-1-Analysen wurde der Zusammenhang von Level, Prompting und Problem-Assoziiiertheit untersucht. Dabei konnte gezeigt werden, dass trotz vielfachen Auftretens von Level-3-Verbalisierungen 80

Session 13 – Mixed Methods (Fr, 20.9., 15:40–17:00, HS B)

.....

Konstruktion universitärer Klausuren auf Basis der Item-Response-Theorie

Frey, Andreas, Born, Sebastian
andreas.frey@uni-jena.de

Prüfungsleistungen werden im Rahmen universitärer Studiengänge häufig mit schriftlichen Klausuren erfasst. Aus methodischer und diagnostischer Sicht sind die eingesetzten Klausuren trotz der hohen Relevanz der Testresultate für die Studierenden oft als defizitär zu bezeichnen. Aus diesem Grund wurde für eine einführende Methodenveranstaltung eine Klausur konstruiert, die (a) die verfolgten Lehrziele adäquat abbilden, (b) kriteriumsorientierte Testwertinterpretationen sowie (c) Unabhängigkeit der Notenkriterien vom Leistungsniveau der jeweiligen Kohorte ermöglichen soll. Als theoretische Rahmenkonzeption diente eine 10×6 -Matrix, in der die Inhaltsbereiche der Vorlesung mit den kognitiven Prozessen nach Bloom et al. (1956) kombiniert und entsprechende kontextualisierte Testaufgaben konstruiert wurden. Die Klausur wurde im Papier- und Bleistift-Format vorgeben und maschinell eingelese. Die Antworten wurden mit dem Programm ConQuest mit dem Partial Credit Modell (Masters, 1982) skaliert. Die Festlegung der Cut-Offs zwischen den Noten erfolgte mit einer vereinfachten Bookmark-Prozedur. Bislang wurde die Klausur zweimal eingesetzt. Um bei beiden Erhebungen die identischen Cut-Off-Werte zu verwenden, erfolgte ein Linking. Die Reliabilität der Skala lag mit .84 und .83 jeweils im hohen Bereich. Im Vortrag wird die Übertragbarkeit des Vorgehens auf andere Inhaltsbereiche sowie die speziellen Anforderungen an Linking und Equating bei universitären Klausuren diskutiert.

Session 1 – Evaluation I (Do, 19.9., 9:20–11:00, HS B)

.....

Markovmodelle und Sequenzanalysen zur Exploration und Modellierung bivariater, kategorialer, wiederkehrender States am Beispiel dyadischer Interaktionsdaten

Fuchs, Peter, Prof. Nussbeck, Fridtjof
Pete.Fuchs@gmail.com

Die Zeit bis zum Auftreten eines bestimmten States (time to event) wird häufig mit Überlebenszeitanalysen modelliert. Dabei können sowohl zeitlich stabile als auch variierende Übergangswahrscheinlichkeiten in das Modell aufgenommen werden; ebenso ist die Modellierung wiederkehrender States (z.B. spezifischer Verhaltensweisen; recurrent events) möglich. Es wird gezeigt, wie bestehende Modellierungsansätze (z.B. expansion of statespace) genutzt werden können, um zwei zeitlich simultan ablaufende State-Prozesse modellieren zu können. Anhand eines Beispiels dyadischer Interaktionsdaten mit 48 Sequenzen wird illustriert, wie das Verhalten eines Partners (state) durch vorhergehendes eigenes Verhalten und das des Partners vorhergesagt werden kann.

Session 16 – IRT VI: Anwendungen / Datenerhebung (Sa, 21.9., 9:30–10:50, HS C)

.....

Parameter mapping between the Diffusion Model and the QUAD Model of IAT data

Gula, Bartosz, Alexandrowicz, Rainer, W.
bartosz.gula@aau.at

The diffusion model has been suggested to capture the distribution of correct and incorrect response times in various speeded decision tasks and typically yields excellent fit to data. In a series of parameter recovery simulations, the boundary separation, drift rate, nondecision components of response time, and variability in the starting point of evidence accumulation were varied in five steps. For each parameter combination 120 single trials were simulated 500 times. It was found that bias in parameter recovery varied depending on the parameter: recovery of boundary separation was nearly unbiased while drift rate and variability in starting point were systematically overestimated. In a second simulation study, it was examined whether the parameters of the diffusion model supposed to represent not controllable and controllable component processes map on conceptually similar parameters of the multinomial QUAD model (Conrey et al., 2005). The QUAD model has been suggested to capture automatic and controlled processes in the Implicit Association Test (IAT). Diffusion model parameters were varied based on fitted values from previous studies (Klauer et al., 2007) to generate plausible data for the IAT. Simulated data was fitted by both models. Preliminary correlations between the fitted parameter values of both models suggest that boundary separation in the diffusion model maps onto both parameters of the QUAD model (D and OB) that reflect controlled processes.

Session 5 – Multinomiale und Diffusionsmodelle (Do, 19.9., 15:20–17:00, HS B)

.....

Vergleich von Auswertungsdesigns quasi-experimenteller Evaluationsstudien: Ergebnisse zum Schulungsprogramm

Hagemann, Anne, Nußbeck, Fridtjof; May, Theodor W.
anne.hagemann@uni-bielefeld.de

Bei der Evaluation von Schulungsprogrammen ist eine randomisierte Zuteilung zu Treatment- und Kontrollgruppe aus praktischen oder ethischen Gründen oft nicht durchführbar. Aufgrund der Selbstselektion können Unterschiede zwischen den Gruppen entstehen, die zu verzerrten Ergebnissen führen. Sowohl Parallelisierungsverfahren als auch die statistische Kontrolle von Gruppenunterschieden können herangezogen werden, um den Grad der Verzerrung zu reduzieren. Within-Study Comparisons zwischen randomisierter und nichtrandomisierter Zuweisung haben gezeigt, dass insbesondere die Auswahl der Kovariaten eine Rolle bei der Reduktion des Bias spielt. Für Schulungsevaluationen, bei denen nur ein quasi-experimentelles Design anwendbar ist und zudem oft ungleiche Gruppengrößen vorkommen, ist jedoch ein Vergleich der Ergebnisse mit auf randomisierter Zuweisung basierenden Befunden nicht möglich. Eine Abschätzung der Verzerrung der Ergebnisse kann der Vergleich verschiedener Auswertungsdesigns innerhalb einer Studie liefern. Am Beispiel der quasi-experimentellen Evaluation eines Schulungsprogrammes für Familien mit epilepsiekranken Kindern wird gezeigt, inwiefern verschiedene Auswertungsdesigns zu unterschiedlichen Ergebnissen gelangen. Folgende Auswertungsstrategien werden verglichen: unabhängige Gruppen ohne statistische Kontrolle von Kovariaten, unabhängige Gruppen mit statistischer Kontrolle von Kovariaten und parallelisierte Gruppen. Auch der Einfluss ungleicher Gruppengröße wird analysiert.

Session 4 – Regression – FA – VA – CA (Do, 19.9., 13:20–15:00, HS C)

.....

IRT-Modelle mit Aufgabenstammabhängigkeiten und Antwortabhängigkeiten in testletbasierten Tests

Hartig, Johannes

hartig@dipf.de

In Schulleistungstests werden häufig Testlets als Testformat eingesetzt, d. h. mehrere Items beziehen sich auf einen gemeinsamen Aufgabenstamm. Dieses Format kann zu lokalen Abhängigkeiten zwischen den Items innerhalb derselben Testlets führen, deren Nichtberücksichtigung bei der Skalierung zu verzerrten Itemparametern und Reliabilitäten führt. Hinsichtlich der Form der Abhängigkeiten innerhalb Testlets kann unterschieden werden zwischen (1) auf den gemeinsamen Aufgabenstamm zurückgehende Abhängigkeiten und (2) spezifischen Effekten des Lösens einzelner Items auf die Lösung von darauf folgenden Items. In Testlet-Modellen werden lokale Abhängigkeiten durch zusätzliche latente Dimensionen für jedes Testlet modelliert. Dieses Vorgehen ist angemessen für die erstgenannte Art von lokalen Abhängigkeiten, Abhängigkeiten des zweiten Typs werden hierbei nicht berücksichtigt. Der Beitrag stellt ein Modell vor, mit dem beide Arten von Abhängigkeiten modelliert werden und unabhängig voneinander geschätzt werden können. Anhand einer Simulationsstudie wird gezeigt, dass beide Effekte prinzipiell unverzerrt getrennt voneinander geschätzt werden können, darüber hinaus werden Mindestanforderungen bezüglich der Datenlage beschrieben. Erste empirische Anwendungen zeigen, dass beide Effekte in empirischen Testdaten beobachtet werden können.

Session 6 – IRT I: Testdesign (Do, 19.9., 15:20–17:00, HS C)

.....

Effekt der Design-Balancierung auf Parameter in IRT-Modellen

Hecht, Martin, Siegle, Thilo; Weirich, Sebastian

martin.hecht@iqb.hu-berlin.de

Beitragsart: Posterlinebreak
In großen Schulleistungstudien kommen häufig Multiple-Matrix-Sampling-Designs zum Einsatz. Hierbei wird jeder Schülerin und jedem Schüler nur ein Teil der gesamten Itemmenge zum Bearbeiten vorgelegt, um den Aufwand für den Einzelnen in Grenzen zu halten. Die Erstellung solcher Designs kann nach verschiedenen Kriterien erfolgen. In einem balancierten unvollständigen Blockdesign wird unter anderem jeder Block mit jedem anderen Block mit gleicher Häufigkeit kombiniert. Zugunsten anderer Randbedingungen (z.B. Begrenzung der Testheftanzahl) kann dieses Kriterium aufgeweicht werden. Dadurch sinkt allerdings die für statistische Auswertungsverfahren zur Verfügung stehende empirische Kovarianz-Information. Das Ziel dieser Studie ist, den Effekt der Design-Balancierung auf Effizienz- und Gütemaße (z.B. Standardschätzfehler und Reliabilität) in IRT-Modellen zu untersuchen. Hierzu wird eine Simulation auf empirischer Datengrundlage durchgeführt. Der zur Verfügung stehende Datensatz besteht aus einer repräsentativen Stichprobe von 25000 deutschen Schülerinnen und Schülern der 9. Klasse, denen 500 Items unter Verwendung eines balancierten Designs vorgelegt wurden. Aus diesem Datensatz werden sukzessive Teilstichproben gezogen, in denen das Design zunehmend weniger balanciert ist. Die aus IRT-Modellen auf diesen Teildatensätzen gewonnenen Effizienz- und Gütemaße werden in Abhängigkeit der Design-Balancierung dargestellt.

Session 6 – IRT I: Testdesign (Do, 19.9., 15:20–17:00, HS C)

.....

Konsequenzen von Modellmisspezifikationen auf die Präzision von individuellen True-Score-Schätzungen

Heene, Moritz, Bühner, Markus; Maraun, Michael
heene@psy.lmu.de

Die Analyse der Auswirkungen von Modellabweichungen im Rahmen von faktorenanalytischen Modellen nimmt in der Psychometrie der letzten 25 Jahre einen großen Raum ein. Klassischerweise wird dabei das Ausmaß der Modellabweichung beispielsweise über die Abweichung zwischen der von den Modellparametern implizierten Kovarianzmatrix und der beobachteten Kovarianzmatrix operationalisiert. Ungeklärt sind bislang allerdings die Auswirkungen von Modellmisspezifikationen in Bezug auf den Kontext für den diese Modelle meist entwickelt und angewandt werden, nämlich der Individualdiagnostik der tatsächlichen Personenmerkmalsausprägung (True-Scores). Das Ziel der vorliegenden Simulationsstudie war es daher, Auswirkungen auf die Präzision von True-Score-Schätzungen unter verschiedenen typischen Modellmisspezifikationen im faktorenanalytischen Modell eines gemeinsamen Faktors zu analysieren. Hierzu wurden unter verschiedenen realitätsnahen Szenarien Daten generiert und mit misspezifizierten Modellen ausgewertet. Die Ergebnisse zeigen, dass die Auswirkungen auch von klassischerweise als gering eingeschätzten Modellmisspezifikationen auf die Präzision der individuellen True-Score-Schätzung praktisch gesehen erheblich sind. Eine formale Erklärung dieser Ergebnisse wird gegeben und die psychometrischen und praktischen Implikationen diskutiert.

Session 3 – SEM I/Pfadmodelle: Anwendung & Caveats (Do, 19.9., 13:20–15:00, HS B)

.....

Die Pairwise-Methode zur Parameterschätzung im ordinalen Rasch-Modell

Heine, Jörg-Henrik, Tarnai, Christian
jhheine@googlemail.com

In sozialwissenschaftlichen Untersuchungen ergeben sich häufig fehlende Werte. Bei mehrstufigen Antwortformaten ergeben sich bei höheren Anteilen fehlender Werte oft sehr geringe Zellohäufigkeiten, die dann zu Problemen bei der Parameterschätzung in probabilistischen Modellen nach der CML- oder MML-Methode führen. Während listwise deletion oder pairwise deletion meist zu einer erheblichen Reduzierung der Stichprobe und so zu verzerrten Schätzungen führt, setzen die meisten Imputationsmethoden a-priori Annahmen über den Daten-Ausfallmechanismus voraus, die meist nicht überprüft werden können. In der vorliegenden Arbeit werden Daten des NEO-PI-R (Items mit 5-stufigem Antwortformat) mit dem R-Paket pairwise analysiert. Die Itemparameter werden dabei nach der Methode des paarweisen Itemvergleichs im Rahmen des Partial Credit Modells explizit berechnet. Diese Methode kann auf Datenmatrizen mit vergleichsweise hohen Anteilen an fehlenden Werten angewendet werden. Die Ergebnisse der Parameterschätzungen nach diesem Verfahren werden mit den Ergebnissen aus der CML-(MML-)Schätzung mit dem Programmpaketen WinMira, eRm und ConQuest verglichen. Zusätzlich wird die Effizienz der Parameterschätzungen nach den drei Methoden (pairwise, CML, MML) bei einem zunehmenden Anteil von fehlenden Werten per Simulation vergleichend dargestellt.

Session 14 – IRT V: Modelltests & Parameterschätzung (Fr, 20.9., 15:40–17:00, HS C)

.....

Zu rekodierende Items: ein Problem für die Skalierung von Persönlichkeitsfragebögen?

Hohensinn, Christine, Khorramdel, Lale
christine.hohensinn@univie.ac.at

Um möglichen Aquieszenztendenzen in Persönlichkeitsfragebogen vorzubeugen, werden häufig innerhalb einer Skala verschieden gepolte Items verwendet. D.h. ein Teil der Items wird in die entgegengesetzte Richtung formuliert und im Auswertungsprozess rekodiert. Implizit wird dabei vorausgesetzt, dass durch die Formulierung in Gegenrichtung dasselbe Konstrukt im Sinne eines bipolaren Merkmals gemessen wird (Marsh, 1996). Außerdem wird bei der Rekodierung der Items die Äquivalenz der Antwortkategorien in vorgesehener und in umgedrehter Reihenfolge vorausgesetzt. Anhand eines mehrdimensionalen Persönlichkeitsfragebogens zur Studieneignung (Wiener Studieneignungs-Persönlichkeitsinventars WSP, Khorramdel & Maurer 2012) wird mittels Methoden der Item Response Theorie (IRT) untersucht, inwieweit die Vorgabe von in die Gegenrichtung formulierten Items als (un)problematisch für die Skalierung betrachtet werden kann. Eindimensionale und multidimensionale IRT Modelle sowie Mixture IRT Modelle werden auf Daten angewendet, die von mehr als 1200 Studierende stammen, welche das WSP im Rahmen eines Self-Assessments bearbeitet haben. Die Ergebnisse werden in Bezug auf die Praxis der Fragebogenkonstruktion und die Interpretierbarkeit von Daten diskutiert.

Session 10 – IRT III: Antwortformate & Itempolung (Fr, 20.9., 11:20–12:40, HS C)

.....

Subskalenhomogenität in der Elternversion des Strengths and Difficulties Questionnaire (SDQ): eine Überprüfung mit probabilistischen Testmodellen

Keller, Ferdinand, Langmeyer, Alexandra
ferdinand.keller@uniklinik-ulm.de

Der SDQ ist ein Instrument zur Messung von psychosozialen Problemen und Stärken von Kindern im Alter von 3–16 Jahren. Er besteht aus 25 Items, mit denen fünf Subskalen gemessen werden. Aufgrund seiner Ökonomie und Validität wird er national und international in zahlreichen klinischen und epidemiologischen Studien eingesetzt. Inzwischen liegen einige Studien zur faktoriellen Struktur sowie zu psychometrischen Eigenschaften im Sinne der klassischen Testtheorie vor, psychometrische Analysen mit probabilistischen Testmodellen blieben bislang die Ausnahme. In der vorliegenden Studie wird der Frage nachgegangen, wie homogen die einzelnen Subskalen sind. Dazu werden Daten des Beziehungs- und Familienpanels pairfam (806 Mütter und 365 Väter) mit Angaben zu ihren 8-15 jährigen Kindern verwendet. Die Analysen erfolgten mit dem Programm WINMIRA 2001 für ordinale Rasch-Modelle sowie LatentGOLD für latent class- bzw- latent factor-Modelle. Die bisherigen Ergebnisse lassen sich zusammenfassend so interpretieren, dass einige Skalen weitgehend homogen i.S. des Rasch-Modells sind (z.B. emotionale Probleme), während z.B. für die Subskala Hyperaktivität eine Lösung mit zwei Faktoren besser geeignet scheint. Den Zusammenhängen mit potentiell konfundierten Variablen, v.a. Alter und Geschlecht, wird derzeit nachgegangen. Weiterhin werden die SDQ-Daten des DJI Surveys AID:A, in der das Altersspektrum der Kinder bei 5-8 Jahren liegt (n=1391), zur Validierung herangezogen.

Session 16 – IRT VI: Anwendungen / Datenerhebung (Sa, 21.9., 9:30–10:50, HS C)

.....

Zentrierung in latenten nichtlinearen Multilevel-Modellen: Eine Monte-Carlo-Studie

Kerwer, Martin, Schermelleh-Engel, Karin und Klein, Andreas G.
martinkerwer@hotmail.de

In Multilevel Strukturgleichungsmodellen (MSEM) können verschiedene nichtlineare Effekte auf latenter Ebene vorliegen, Effekte auf Level 1 oder 2 sowie Cross-Level-Interaktionen. Ein Problem nichtlinearer Modelle besteht in der Multikollinearität zwischen Prädiktoren und Produkttermen. Bei mehreren nichtlinearen Effekten bewirkt die Korrelation der Prädiktoren, dass nichtlineare Terme korrelieren und diese Effekte schwer zu differenzieren sind. Um Multikollinearität zu reduzieren, ist in MSEM eine Zentrierung am Gesamtmittelwert oder Gruppenmittelwert durchführbar. Unklar ist, welche Zentrierung in den unterschiedlichen nichtlinearen Modellen zu den zuverlässigsten Ergebnissen führt. Mit der vorliegenden Studie soll untersucht werden, ob verschiedene Arten der Nichtlinearität in MSEM differenzierbar sind, ob bei Fehlspezifikationen fälschlicherweise nichtlineare Effekte aufgedeckt werden und wie sich die Art der Zentrierung auf Schätzgüte und Interpretation nichtlinearer Parameter auswirkt. In einer Monte-Carlo-Studie wurden Nichtlinearität, Fehlspezifikation der Modelle, Zentrierung, Korrelation zwischen den Prädiktoren, ICC, Clustergröße und Clusteranzahl variiert. Die Daten wurden in R generiert und in Mplus mit dem Produktindikatoransatz analysiert. Erste Ergebnisse zeigen, dass die verschiedenen nichtlinearen Effekten zufriedenstellend identifiziert werden können. Die Konsequenzen der Art der Zentrierung werden diskutiert.

Session 9 – Mehrebenenmodelle (Fr, 20.9., 11:20–12:40, HS B)

.....

Erfassung und Korrektur von Antworttendenzen in Large Scale Assessments mittels eines neuen Multidimensionalen Item Response Theorie Ansatzes

Khorramdel, Lale, von Davier, Matthias, Yamamoto, Kentaro
lkhorramdel@ets.org

Vorgestellt wird ein Multidimensionaler Item Response Theorie (MIRT) Ansatz zur Erfassung und Korrektur von Antworttendenzen in Ratingskalen, welche die Ergebnisse wissenschaftlicher Untersuchungen verfälschen und die Validität von Messinstrumenten vermindern können (vgl. De Jong, Steenkamp, Fox & Baumgartner, 2008; Dolnicar & Grun, 2009; Weijters, Schillewaert & Geuens, 2008). Basierend auf dem Ansatz von Böckenholt (2012) und Meiser und Böckenholt (2011) werden die Antworten auf einer 5-stufigen Ratingskala in einzelne aufeinanderfolgende Antwortprozesse zerlegt und mit MIRT Modellen modelliert. Untersucht werden die Antworttendenzen “Tendenz zu Extremantworten” und “Tendenz zur Mitte”. Die Daten stammen von einem Hintergrundfragebogen einer internationalen Large Scale Assessment Studie, welcher verschiedene Skalen erfasst, von denen angenommen wird, dass sie mit der Leistung im kognitiven Teil des Assessments zusammenhängen. Es wird aufgezeigt, dass Antworttendenzen eindimensional erfasst und von Konstrukt-relevanten Antworten differenziert werden können. Die Ergebnisse werden im Hinblick auf mögliche kulturelle Unterschiede diskutiert und der Einfluss von Antworttendenzen auf die Untersuchung des Zusammenhangs zwischen Hintergrundvariablen und kognitiven Testleistungen wird vorgestellt.

Session 12 – IRT IV: Antwortverhalten & Antwortstile (Fr, 20.9., 14:00–15:20, HS C)

.....

Fallacies and misconceptions when modeling mediation

Klein, Andreas, Brandt, Holger
klein@psych.uni-frankfurt.de

In many research contexts in the social or health sciences, researchers are interested in identifying intermediate variables that stand in the pathway or may serve as early indicators of a developing disease or critical behavior. Causal effects that are indirect (mediated) and go through intermediate variables that act as components in a causal chain are also known as mediator effects. In the biostatistical literature, however, serious concerns about the conventional methodology such as path analysis used to explore those effects have been raised. Following this critique, a causal interpretation of a path-analytic mediator effect is hardly ever justified, and it can be shown that path analysis may often identify supposedly strong mediator effects that in reality are nothing but methodological artefacts. In this paper, we explain why the standard mediator model is conceptually flawed, and why it requires additional, untestable assumptions to become interpretable as a model that can represent a mediated causal process. Directions for future model development and consequences for statistical applications are discussed.

Session 3 – SEM I/Pfadmodelle: Anwendung & Caveats (Do, 19.9., 13:20–15:00, HS B)

.....

Berücksichtigung der Tendenz zum Überspringen von Aufgaben bei der Schätzung von Kompetenzscores

Köhler, Carmen, Pohl, Steffi; Carstensen, Claus H.
carmen.koehler@uni-bamberg.de

Bei der Erfassung von Fähigkeiten mithilfe von Kompetenztests überspringen Personen häufig einzelne Aufgaben. Neuere modell-basierte Ansätze nehmen neben der Fähigkeit die missing propensity als zweite latente Variable in das Messmodell auf, wobei zum Schätzen der Parameter mithilfe der marginalen Maximum-Likelihood (MML)-Methode eine bivariate Normalverteilung der Fähigkeit und der missing propensity angenommen wird. Dies stellt nur dann eine angemessene Methode zur Berücksichtigung nicht ignorierbarer fehlender Werte dar, wenn ein Persönlichkeitsmerkmal, welches das Überspringen einzelner Items bestimmt, existiert und den modell-basierten Ansätzen entsprechend skalierbar ist. Unter Verwendung der Daten des Nationalen Bildungspanels (N = 15 396) werden missing propensities in vier Kompetenztests Rasch-modelliert und auf dessen Passung geprüft. In der Regel ist für die missing propensity eine Normalverteilung nicht gegeben, weshalb die Robustheit der Parameterschätzer auch bei Verletzung dieser Annahme untersucht wird. Hierzu dienen Vergleiche zwischen Modellen mit und ohne Normalverteilungsannahme sowie einem HYBRID Modell, welches die tatsächliche Verteilung in den Daten angemessen beschreibt. Die Ergebnisse lassen darauf schließen, dass die missing propensities entsprechend der modell-basierten Ansätze modellierbar sind und diese sich somit dafür eignen, nicht ignorierbare fehlende Werte bei der Skalierung angemessen zu berücksichtigen.

Session 12 – IRT IV: Antwortverhalten & Antwortstile (Fr, 20.9., 14:00–15:20, HS C)

.....

Quasi-exakte Tests für das Rasch Modell: Anwendung als Voraussetzungsüberprüfung für Veränderungsmessungsstudien

Koller, Ingrid, Wiedermann, Wolfgang, Glück, Judith
ingrid.koller@univie.ac.at

Das Rasch Modell (Rasch, 1960) zur Überprüfung der Item-Qualität zeichnet sich durch zahlreiche positive Eigenschaften (z. B. spezifische Objektivität) aus. Eine Möglichkeit der Modellüberprüfung sind quasi-exakte Tests (z. B. Koller & Hatzinger, 2013; Koller, Alexandrowicz & Hatzinger, 2012; Ponocny, 2001) die sich besonders zur Überprüfung der Item-Qualität bei kleinem Stichprobenumfang eignen. Diese Teststatistiken eignen sich jedoch nicht nur für die Überprüfung des Rasch Modells, sondern bieten weitere Anwendungsmöglichkeiten. Im Kontext der Veränderungsmessung kann untersucht werden, ob sich Items zu unterschiedlichen Zeitpunkten homogen und lokal stochastisch unabhängig zueinander verhalten und somit keine interindividuellen Unterschiede hinsichtlich der Veränderung innerhalb von Personengruppen vorliegen. Der Vortrag gibt zunächst einen allgemeinen Überblick über quasi-exakte Tests zur Prüfung des Rasch Modells. Des Weiteren werden Anwendungen im Falle longitudinaler Daten diskutiert.

Session 14 – IRT V: Modelltests & Parameterschätzung (Fr, 20.9., 15:40–17:00, HS C)

.....

Neue Entwicklungen der Mehrebenenmodellierung bei ambulanten Daten

Krenn, Sabine, Schwerdtfeger, Andreas
sabine.krenn@uni-graz.at

Bei psychophysiologischen Fragestellungen werden Längsschnittdaten vermehrt mit Mehrebenenmodellen analysiert, da diese die individuelle Variation in den zeitlichen Verläufen berücksichtigen und aufklären. Kuljanin, Braun und DeShon (2011) konnten zeigen, dass die Mehrebenenmodellierung allerdings nur bei rein deterministischen Prozessen zuverlässige Ergebnisse liefert und bei stochastischen Prozessen das Random-Coefficient-Modell fälschlicherweise angenommen wird. Sie schlagen daher eine Analyse mittels Unit-Root-Test vor, um die Art des datengenerierenden Prozesses zu überprüfen. Insbesondere bei kardialen Daten ist diese Art des zugrundeliegenden Prozesses umstritten. Daher untersuchte die vorliegende Studie EKG-Daten (Herzrate), die an 122 Personen über 72 Stunden mittels der Methodik des ambulanten Assessments erhoben wurden. Die kardialen Daten wurden mit dem Panel-ADF-Test auf interindividueller Ebene und mit dem ADF-Test sowie dem KPSS-Test auf den intraindividuellen Ebenen analysiert. Die Ergebnisse der gesamten Analysen können einen deterministischen datengenerierenden Prozess der kardialen Parameter nicht hinreichend nachweisen, da die Verfahren nur eine geringe Teststärke besitzen. Aufgrund dieser Ergebnisse erweisen sich Zeitreihenanalysen gegenüber Mehrebenenmodellierung für kardiale Parameter als geeigneteres Verfahren.

Session 9 – Mehrebenenmodelle (Fr, 20.9., 11:20–12:40, HS B)

.....

Unterschiedliche Skalennutzung bei der Messung von Arbeitszufriedenheit: Anwendung des gemischten generalisierten Partial-Credit-Modells

Kutscher, Tanja, Crayen, Claudia; Eid, Michael
 tanja.kutscher@fu-berlin.de

Arbeitszufriedenheit gilt in der Sozialforschung als Standardindikator für Lebenszufriedenheit. Die Erfassung in Panelstudien erfolgt meist anhand weniger Items mit vielen Antwortkategorien. So wurden z. B. im “Household, Income and Labour Dynamics in Australia” (HILDA) Panel fünf Aspekte der Arbeitszufriedenheit mit jeweils einer 11-stufigen Antwortskala erhoben. Diese feine Abstufung kann zur Überforderung der Befragten und unterschiedlicher Skalennutzung führen. Um die Heterogenität der Skalennutzung zu untersuchen, wurden auf Daten des HILDA-Panels (Welle von 2001, n=7036) das gemischte generalisierte Partial-Credit-Modell (PCM) angewendet. Mit diesem Modell lassen sich latente Klassen mit unterschiedlicher Skalennutzung identifizieren und Personwerte auf der latenten Dimension schätzen. Im Gegensatz zum einfachen PCM werden im generalisierten PCM die Diskriminationsparameter der Items frei geschätzt. Anhand von Informationskriterien wurde ein Modell mit drei Klassen ausgewählt. Keine der gefundenen Klassen nutzt alle der elf Antwortkategorien. Klassenunterschiede betreffen etwa die Anzahl der genutzten Antwortkategorien und die Präferenz für bestimmte Skalenbereiche. Die Vorhersage der Klassenzugehörigkeit erfolgt durch soziodemografische, Persönlichkeits- und arbeitskontextbezogene Variablen. Die Angemessenheit von den verwendeten Items mit vielen Antwortkategorien zur Erfassung der Arbeitszufriedenheit wird zur Diskussion gestellt.

Session 12 – IRT IV: Antwortverhalten & Antwortstile (Fr, 20.9., 14:00–15:20, HS C)

.....

Das Rasch-Modell für Trauma-Inventare: Ein Anwendungsbeispiel zur Dichotomisierung des Harvard Trauma Questionnaire bei kleinen Stichproben.

Leuschner, Marcel, Mochar, Elena Maria; Alexandrowicz, Rainer, W.
 marcel.leuschner@aau.at

Das dichotome Item-Response Modell nach Rasch (1960) ist vornehmlich für seine Anwendung in der Leistungstestung bekannt. Als Ergänzung zu einem qualitativen Urteil ist jedoch auch eine Ausweitung probabilistischer Testtheorie auf bekannte Trauma-Inventare vorstellbar. In einem Anwendungsbeispiel für eine kleine Stichprobe (n=50) in Kärnten lebender, afghanischer Asylwerber wurden Globalwerte des Harvard Trauma Questionnaire (HTQ) anhand zweier Teilungskriterien gesplittet. Zum einen, ein gegebener Cut-Off-Wert aus der Literatur sowie zum anderen, der Grenzwert einer eigens hierfür durchgeführten Receiver Operating Characteristic (ROC)-Analyse. Die anschließende Diskussion der Raschkonformität (GOF) stützte sich aufgrund des geringen Stichprobenumfangs auf nichtparametrischer Verfahren (Koller et al., 2012) im R-Paket *eRm* (Mair & Hatzinger, 1997), anstatt des Rückgriffs auf die parametrische Variante zur Modelltestung durch den Likelihood Ratio Test (Andersen, 1973). Anhand des Anwendungsfalles wird der Nutzen der verwendeten Verfahren diskutiert und eine praktische Erweiterung zur Beurteilung des vorgestellten Trauma Screening-Instruments gegeben.

Session 16 – IRT VI: Anwendungen / Datenerhebung (Sa, 21.9., 9:30–10:50, HS C)

.....

Verlässlichkeit der WLS-Fitstatistiken bei der Schätzung eindimensionaler Faktorenmodelle mit ordinalen Daten

Lubbe, Dirk, Schuster, Christof

Dirk.Lubbe@psychol.uni-giessen.de

Weighted least squares ist eine ökonomische Methode zur Schätzung von Faktormodellen aufgrund kategorialer Indikatoren. Untersuchungen und Anwendungsempfehlungen dieser Methode wurden meist auf geringe Anzahlen von Variablen begrenzt, weil die Berechnung der benötigten Gewichtungsmatrix für viele Indikatoren oft problematisch ist. Bei Verwendung einer diagonalen Gewichtungsmatrix können höhere Anzahlen von Variablen berücksichtigt werden, sodass das Verfahren im Zusammenhang mit der Konstruktion psychologischer Test angewendet werden kann. Um die Güte der Modellanpassung beurteilen zu können, bedarf es allerdings geeigneter Fit-Statistiken. Eine Möglichkeit hierfür sind robuste Chi-Quadrat-Statistiken nach Satorra und Bentler. In Simulationsstudien haben wir Schätzungen für verschiedene Variablen- und Kategorienzahlen sowie verschiedene Stichprobenumfänge durchgeführt. Bei Schätzung mit vollständiger Gewichtungsmatrix resultiert bei zunehmender Variablenzahl schnell eine überproportionale Ablehnung des Modells. Bei Schätzung mit diagonaler Gewichtungsmatrix zeigen die Satorra-Bentler-Statistiken hingegen eine verlässlichere Annäherung an das nominale Alpha-Niveau.

Session 3 – SEM I/Pfadmodelle: Anwendung & Caveats (Do, 19.9., 13:20–15:00, HS B)

.....

Dirichlet-Multinomiale Regressionsmodelle in der Psychologie

Maier, Marco J.

marco.maier@wu.ac.at

Ein weitverbreiteter Datentyp in den Sozialwissenschaften sowie der Psychologie sind kategoriale Daten. Ordinale und nominale abhängige Variablen werden üblicherweise mit (ordinalen, bzw. multinomialen) logistischen Regressionsmodellen ausgewertet. Um diese Verfahren möglichst einfach zu halten wird die Anzahl der Durchgänge in der Praxis meist auf $n = 1$ (d.h. nur eine Kategorie kann gewählt werden) eingeschränkt. Vor allem bei nominalen Daten (sprich ≥ 2 Kategorien ohne inhärente Ordnung) vereinfacht dieses Vorgehen die Analyse, kann aber auch zu erheblichem Informationsverlust führen, da Versuchspersonen gezwungen sind, sich für lediglich eine der Alternativen zu entscheiden. Die multinomiale logistische Regression ist ein gutes Modell für Daten mit $n = 1$, jedoch führen $n > 1$ Durchgänge möglicherweise zu Phänomenen wie "overdispersion", da die Multinomialverteilung keine Möglichkeit besitzt, Streuung zu modellieren. Als Alternative wird eine sog. "Compound-Verteilung" aus Dirichlet- und Multinomialverteilung vorgestellt, die weitaus flexibler ist und, in geeigneter Parametrisierung, eine explizite Schätzung und Modellierung der Streuung erlaubt. Die Vorteile und Anwendung dieses Zugangs werden mit Beispielen und Analysen in R illustriert.

Session 4 – Regression – FA – VA – CA (Do, 19.9., 13:20–15:00, HS C)

.....

Parameter recovery der Modellparameter in Diffusionsmodellen

Martin, Silke, Alexandrowicz, Rainer, W.; Gula, Bartosz

silke.schaefer@edu.uni-klu.ac.at

Das Diffusionsmodell von Ratcliff (1978) modelliert die Korrektheit und Reaktionszeit von Antworten in binären Entscheidungsaufgaben als einen Wiener Prozess mit sieben Parametern und wird in zunehmend mehr Anwendungsbereichen dazu verwendet, vier latente Komponenten kognitiver Prozesse zu entmischen. Zur Auswertung stehen mehrere Software-Lösungen zur Verfügung, u.A. Fast-dm (Voss & Voss, 2007), das R-Package Rwiener (Wabersich, in Vorbereitung) und DMAT (Vandekerckhove & Tuerlinckx, 2007). In einer Simulationsstudie wird für die drei Programme die Äquivalenz von Modellfit und Parameterschätzungen überprüft.

Session 5 – Multinomiale und Diffusionsmodelle (Do, 19.9., 15:20–17:00, HS B)

.....

Multinomiale Verarbeitungsbaum-Modelle zur Erfassung latenter kognitiver Prozesse bei Shooter-Aufgaben

Mayerl, Hannes

hannes.mayerl@aau.at

Shooter-Aufgaben werden häufig eingesetzt, um den Einfluss von Stereotypen auf das Verhalten zu erfassen. Solche Aufgaben sind einem Computerspiel ähnlich, bei denen bewaffnete Personen erschossen und unbewaffnete Personen am Leben gelassen werden sollen. In Studien an amerikanischen Stichproben konnte gefunden werden, dass unbewaffnete Schwarzafrikaner im Vergleich zu unbewaffneten weißen Personen häufiger fälschlicherweise erschossen werden. In der vorliegenden Studie wurde untersucht, ob ein solcher Effekt auch bei einer europäischen Stichprobe zu finden ist. Dafür mussten die Studierenden einer österreichischen Universität ($N = 41$) eine ähnliche Shooter-Aufgabe wie bei Correll, Park, Judd & Wittenbrink (2002) durchführen. Zusätzlich wurden die Studierenden gebeten, einen Zeitungsartikel über schwarzafrikanische (vs. westeuropäische) Kriminelle zu lesen, um die stereotype Assoziation zwischen Schwarzafrikanern und Gefahr zu aktivieren. Die europäischen Versuchspersonen weisen in der Shooter-Aufgabe ein ähnlich stereotypes Antwortverhalten auf wie die amerikanischen Versuchspersonen. Ein Vergleich zwischen drei multinomialen Verarbeitungsbaum-Modellen zeigt, dass das Prozess-Dissoziations-Modell (PD-Modell) am besten geeignet ist, die Daten aus dieser Aufgabe zu beschreiben. Weitere Analysen mit dem PD-Modell legen nahe, dass das Priming mit dem Zeitungsartikel einen Einfluss auf die automatischen (A), jedoch nicht auf die kontrollierten Prozesse (C), ausgeübt hat.

Session 5 – Multinomiale und Diffusionsmodelle (Do, 19.9., 15:20–17:00, HS B)

.....

Analyse von Effekten der Itempolung mit ordinalen IRT-Modellen: Trait-Struktur und Antworttendenz

Meiser, Thorsten, Fleig, Hanna

t.meiser@uni-mannheim.de

Zur Analyse von Fragebogendaten mit positiv formulierten und invertierten Items wurden Strukturgleichungsmodelle vorgeschlagen, die neben den Traits der zu messenden Konstrukte Methodenfaktoren zur Berücksichtigung individueller Effekte der Itempolung spezifizieren (DiStefano & Motl, 2009; Pohl & Steyer, 2010). Neben den beobachteten Effekten auf die Trait-Struktur kann sich die Itempolung jedoch auch auf die Nutzung des Antwortformats in Rating-Items auswirken, was durch ordinale IRT-Modelle abgebildet werden kann (Hinz, Brähler, Geyer & Körner, 2003; Meiser, Hoeffler & Jasper, 2012). Daher werden die vorgeschlagenen Faktorenmodelle in diesem Vortrag durch geeignete Link-Funktionen zu mehrdimensionalen ordinalen IRT-Modellen für Rating-Daten mit invertierten Items generalisiert und auf reale Daten angewendet. Die Modellspezifikation erfolgt durch hierarchische Modelle, die eine flexible Festlegung der Trait-Struktur und der Schwellenparameter ordinaler IRT-Modelle erlauben sowie die Einbeziehung von Kovariaten ermöglichen. Mit einem Datensatz von $N=1024$ Respondenten zum Konstrukt Personal Need for Structure wird die Spezifikation inhaltlicher Traits und eines Methodenfaktors mit unterschiedlichen Modellannahmen über die Nutzung des Antwortformats in Abhängigkeit der Itempolung illustriert. Die Ergebnisse zeigen, dass individuelle Effekte der Itempolung als Suppressor von Kovariaten fungieren können und dass die Itempolung die Nutzung des Antwortformats beeinflusst.

Session 10 – IRT III: Antwortformate & Itempolung (Fr, 20.9., 11:20–12:40, HS C)

.....

Lässt sich Emotionsregulierung mittels fMRI nachweisen?

Morawetz, Carmen, Alexandrowicz, Rainer W., Heekeren, Hauke R.
carmen.morawetz@fu-berlin.de

Neuroimaging studies in cognitive neuroscience have traditionally relied on group analyses that attempt to characterize common activations across subjects and that usually regard variance among individuals as statistical noise. However, in the domain of emotion processing, individual differences in responses are the rule rather than the exception. In light of these individual differences, a growing number of studies have adopted the useful complementary approach of including measures of individual differences in the analysis to uncover brain areas whose activity covaries with these measures rather than relying solely on group-averaged brain activations. There is rising interest in the study of the neural correlates of such individual differences in emotion regulation given that individuals who tend to use reappraisal more often report experiencing less negative affect, better social functioning and greater psychological well-being. In the present study we aim to independently assess personality traits in a large sample and directly examine their relationship to emotion regulation success and its underlying neuronal network using an explicit emotion regulation task. We use structural equation modeling to investigate the mediational role of various personality traits on emotion regulation success and its neuronal basis bridging the gap between personality psychology and neuroscience. The contribution will present requirements, strengths and limitations of the SEM approach to model fMRI data.

Session 2 – Prozessmonitoring & fMRI (Do, 19.9., 9:20–11:00, HS C)

.....

Statistische Modelle für Hochschul-Rankings am Beispiel der Psychologie

Mutz, Rüdiger, Daniel, Hans-Dieter
mutz@gess.ethz.ch

Während Rankings von Hochschulen nicht nur für den Bereich Forschung (z.B. Leiden Ranking) sondern für den Bereich Lehre und Studium (z.B. CHE Hochschulranking) wachsende Bedeutung in Politik, Wissenschaft und Gesellschaft gewinnen, sind die methodisch-statistischen Grundlagen dieser Rankings meistens unbefriedigend und spiegeln nicht unbedingt den gegenwärtigen methodisch-statistischen Forschungsstand wider. Im Hinblick auf Rankings in Bereich Lehre und Studium sollen anhand zweier Befragungsbatterien zur Studienortwahl und zu Studienbedingungen, neuere statistische Modelle von Rankings dargestellt werden: klassische Mehrebenenanalyse-Modelle, Logit-Modelle von Mengen rang-geordneter Objekte, Modelle mit Berücksichtigung von Kovariablen, Panel-Modelle, Missing-Imputations-Modelle unter Berücksichtigung von Vorinformationen und Bayesianische Modelle. Im Rahmen einer Sekundäranalyse von Daten einer Mehrthemen-Panel-Befragung aller Studierenden (N = 1490), die im Wintersemester 1999/2000 ihr Diplom-Studium im Fach "Psychologie" in Deutschland aufgenommen haben und 2007 erneut befragt wurden, soll eine Auswahl von Modelle exemplarisch angewendet und auf ihre Geeignetheit für praktischen Anwendungen geprüft werden.

Session 1 – Evaluation I (Do, 19.9., 9:20–11:00, HS B)

.....

Wie kann der Studienerfolg im Fach Psychologie erhöht werden? Konzeption und Evaluation von Maßnahmen zur Förderung mathematischer Kompetenzen

Niemeier, Sarah, Nussbeck, Fridtjof
sarah.niemeier@uni-bielefeld.de

Prädiktoren für den Studienerfolg im Fach Psychologie sind neben Schulnoten, Intelligenz und Persönlichkeit vor allem mathematische Kompetenzen. Studienerfolg wird dabei nicht nur über Studiennoten und Studiendauer sondern auch über einen erfolgten Studienabschluss/ Studienabbruch und Studienzufriedenheit operationalisiert. An der Universität Bielefeld wurden im WS 2012/2103 und im SoSe 2013 Lehr- und Beratungsmaßnahmen zur Förderung mathematischer Kompetenzen speziell im ersten Studienjahr konzipiert, um die Lehrqualität zu steigern und die Studienbedingungen zu verbessern sowie langfristig die Studienabbruchsquote zu verringern und den Studienerfolg zu steigern. Bei der Entwicklung und Umsetzung der Maßnahmen wurden Konzepte der Mathematikdidaktik berücksichtigt. Die Maßnahmen (u.a. Online-Fragebogen zur Vorlesung Statistik, Beratungssprechstunde, Arbeitsblätter, digitale Lernmodule, zusätzliches Tutorium) wurden in einer längsschnittlichen Erhebung evaluiert, indem Häufigkeit und Intensität der Nutzung mit Maßen für Studienerfolg (Statistiknote, Verständnis relevanter Inhalte und Studienzufriedenheit) in Beziehung gesetzt wurden. Erste Ergebnisse, Ansätze zur Optimierung der Maßnahmen sowie Implikationen für das Lehrangebot im ersten Studienjahr werden vorgestellt und diskutiert.

Session 15 – Evaluation II (Sa, 21.9., 9:30–10:50, HS B)

.....

Das Common Fate und das Actor Partner Interdependence Modell zur Analyse dyadischer Daten

Nussbeck, Fridtjof W.
fridtjof.nussbeck@uni-bielefeld.de

Im Rahmen der Analyse dyadischer Daten werden unter anderem das Common Fate und das Actor Partner Interdependence Modell eingesetzt. Beide Modelle werden eingesetzt, um regressive Beziehungen zu modellieren. Dem Common Fate Modell liegt die Annahme zugrunde, dass Mitglieder einer Dyade (Partner) ein gemeinsames Schicksal teilen, so dass eine Kriteriums- und eine Prädiktorvariable in die regressive Beziehung aufgenommen werden. Dem Actor Partner Interdependence Modell liegt die Annahme zugrunde, dass Partner sich zwar gegenseitig beeinflussen, dies jedoch nicht perfekt, so dass eine Kriteriums- und eine Prädiktorvariable für jeden der Partner in das Modell aufgenommen werden muss. Die Kriterien werden mit den Prädiktoren beider Partner vorhergesagt. Obwohl sich die Modelle konzeptuell unterscheiden und zur Beantwortung unterschiedlicher Forschungsfragen herangezogen werden, führen sie unter bestimmten Spezifikationen zu datenäquivalenten Modellen und sind somit statistisch nicht voneinander zu trennen. Die Ansprüche an die Modellspezifikationen, die den genannten Annahmen entsprechen, werden abgeleitet und anhand eines empirischen Beispiels illustriert.

Session 3 – SEM I/Pfadmodelle: Anwendung & Caveats (Do, 19.9., 13:20–15:00, HS B)

.....

Prognosegüte von Auswahlverfahren: Ein Modell zur Korrektur der Kriteriumsvalidität bei dichotomer Erfolgsbewertung

Pfaffel, Andreas, Schober, Barbara, Spiel, Christiane
andreas.pfaffel@univie.ac.at

Ein zentrales und lang bekanntes Problem bei der Prüfung der Kriteriumsvalidität von Auswahlverfahren (oft in der Form als Eignungstests) ist, dass die Daten systematisch zensiert sind und damit eine range restriction vorliegt. Nur KandidatInnen, die als geeignet identifiziert und ausgewählt wurden, können hinsichtlich des Erfolgskriteriums beurteilt werden. Unter den Nicht-Ausgewählten ist der Anteil der Falsch-Klassifizierten nicht bekannt. Dies führt zu verzerrten Schätzern. Korrekturmethode bei intervallskalierten Erfolgskriterien sind entwickelt und mehrfach geprüft, eine zufriedenstellende Korrekturmethode für dichotome Erfolgskriterien, welche jedoch häufig in realen Situationen anzutreffen sind (z.B. im Hochschulbereich: Studium abgeschlossen vs. nicht abgeschlossen), liegt bisher jedoch noch nicht vor. Auf Basis einer zweiparametrischen logistischen Regression wurde ein Modell entwickelt, das trotz zensierter Daten eine statistische Schätzung der Anzahl nicht ausgewählter KandidatInnen, die das Kriterium jedoch erreicht hätten, ermöglicht. Dadurch kann die Anzahl korrekter Klassifizierungen (die Trefferquote) quantifiziert werden. Die Schätzung ist ohne Einbezug weiterer Daten möglich. Einflussfaktoren auf die Güte der Schätzung wie Selektionsrate, Höhe der Korrelation zwischen Auswahl- und Erfolgskriterium sowie die Basisrate des Erfolgs wurden mittels Monte-Carlo-Simulation systematisch geprüft. Möglichkeiten und Grenzen des Verfahrens werden diskutiert.

Session 11 – Wachstumskurven & GMM (Fr, 20.9., 14:00–15:20, HS B)

.....

Überprüfung der dimensionalen Struktur einer Adaptation des Trierer Inventars zur Lehrevaluation (TRIL)

Pfeiffer, Andreas, Fondel, Eva; Weis, Susanne; Lischetzke, Tanja; Gollwitzer, Mario
pfeiffer@uni-landau.de

Eine weit verbreitete Methode zur Evaluation von Lehrveranstaltungen und der Erfassung des Konstrukts "Qualität der Lehre" ist der Einsatz von standardisierten Fragebögen wie z.B. dem Trierer Inventar zur Lehrevaluation (TRIL; vgl. Gollwitzer, Kranz & Vogel, 2006; Gollwitzer & Schlotz, 2003). Die Evaluation mittels solcher Instrumente gewährleistet Vergleichsmöglichkeiten zwischen und innerhalb von Lehrveranstaltungen über die Zeit und bietet darüber hinaus Lehrenden Rückmeldung über Verbesserungspotentiale ihrer Lehre. Zudem liefern die gewonnenen Daten die Grundlage um Lehrpreise und Weiterbildungsangebote für Lehrende zu implementieren. An der Universität Koblenz-Landau wird zur Evaluation der Lehre eine adaptierte Variante des TRIL eingesetzt, deren Nützlichkeit und Validität in vorangegangenen Studien untersucht wurde. Exploratorische Faktorenanalysen legten eine Reduktion auf die vier Dimensionen Struktur & Didaktik, Anregungsgehalt, Veranstaltungsklima und Anwendungs- & Praxisbezug nahe. In dieser Studie wurde diese vorgeschlagene Struktur anhand von $N = 181$ Veranstaltungen durch eine konfirmatorische Faktorenanalyse, die die hierarchische Struktur der Daten berücksichtigt, untersucht. Da die multivariate Normalverteilung der Indikatoritems nicht angenommen werden kann, wurden diese als ordinale Variablen in das Modell eingefügt und ein WLSMV-Schätzer genutzt. Die Passung der Struktur, sowie die Messinvarianz über verschiedene Semester und Fachbereiche werden diskutiert.

Session 15 – Evaluation II (Sa, 21.9., 9:30–10:50, HS B)

.....

Validierung eines neuen Modells zur Trennung von Traitausprägung und Antwortstilen

Plieninger, Hansjörg, Meiser, Thorsten
 plieninger@uni-mannheim.de

Die Tendenz Fragebogenitems (teilweise) unabhängig vom Iteminhalt zu beantworten wird allgemein als Antwortstil (response style) bezeichnet und gefährdet die Gütekriterien und die Interpretation entsprechender Instrumente. Kürzlich stellte Böckenholt (2012) ein Prozess-Modell vor, das es erlaubt Ratingdaten in Antwortstile auf der einen Seite und die eigentliche, inhaltliche Traitausprägung auf der anderen Seite zu zerlegen. Das Modell – ein mehrdimensionales IRT-Modell – basiert auf der Rekodierung der ursprünglichen Antworten in sog. Pseudoitems, die so gewählt werden, dass sie entweder mit dem Iteminhalt oder einem Antwortstil korrespondieren. Ziel der vorliegenden Studie war es die Annahmen des Prozess-Modells anhand externer Kriterien zu validieren. Im Rahmen des Konstanzer Studierendensurveys (Ramm, Multrus & Bargel, 2011) beantwortete eine Stichprobe von Studierenden (N = 7483) neun 7-stufige Items aus dem Bereich studentisches Selbstvertrauen. Ein Prozess-Modell, das neben der inhaltlichen Traitdimension zusätzliche Dimensionen von Antwortstilen berücksichtigte, war einem eindimensionalen Modell überlegen. Als externe Kriterien dienten einerseits die Studiennote und andererseits Indikatoren für Antwortstile basierend auf 60 nahezu unkorrelierten Items. Es zeigten sich hypothesenkonforme Assoziationen und Dissoziationen zwischen den Kriterien und den verschiedenen Dimensionen des Prozess-Modells.

Session 12 – IRT IV: Antwortverhalten & Antwortstile (Fr, 20.9., 14:00–15:20, HS C)

.....

Mixed-Rasch-Hybrid Modelle zur Identifikation von testbaren Förderschülern

Pohl, Steffi, Hardt, Katinka; Südkamp, Anna; Nusser, Lena; Weinert, Sabine
 steffi.pohl@uni-bamberg.de

Large-scale assessments wie das Nationale Bildungspanel (NEPS) versuchen Förderschüler in die Kompetenzmessung zu integrieren. Die Herausforderung dabei ist Kompetenztests so zu entwickeln, dass diese auch bei Förderschülern zu reliablen und zu Regelschülern vergleichbaren Messungen führen. Bisherige Machbarkeitsstudien im NEPS, die neben Standardtests für Regelschüler auch in Schwierigkeit und Länge angepasste Tests implementiert haben, zeigen, dass dies kein leichtes Unterfangen ist. In diesem Vortrag widmen wir uns nicht der Frage danach wie ein Test konstruiert sein muss, damit er bei Förderschülern zu reliablen und vergleichbaren Messungen führt, sondern mit welchen Modellen man Förderschüler identifizieren kann für die man mit Tests aus large-scale assessments adäquate Kompetenzmessungen erhält. Dazu wurden Mixed-Rasch-Hybrid-Modelle spezifiziert, die unterschiedliche Bearbeitungsstrategien von Schülern abbilden. Die Bearbeitungsstrategien werden mit Hilfe von Informationen zum Auftreten von fehlenden Werten sowie durch die Modellierung von Ratewahrscheinlichkeiten abgeleitet. Die Analyseergebnisse von Daten aus Machbarkeitsstudien des Nationalen Bildungspanels von Förderschülern mit Lernbehinderungen der Klassenstufe 9 zeigen, dass es mit den Modellen möglich ist eine Gruppe von Förderschülern zu identifizieren für die das Rasch-Modell passt und die Messinvarianz mit Regelschülern aufweisen. Diese Gruppe kann von Förderschulgruppen abgegrenzt werden, die die richtige Lösung der Testaufgaben raten oder Aufgaben nur teilweise bearbeiten. Es wird versucht die identifizierten Gruppen durch erklärende Variablen zu beschreiben, die es in weiteren Studien ermöglichen bereits vor der Testung Förderschüler auszuwählen, deren Kompetenz mit den vorgegebenen Tests gut erfasst werden kann.

Session 8 – IRT II: Modellentwicklung (Fr, 20.9., 10:20–11:00, HS C)

.....

Die Erfassung von subjektivem Wohlbefinden mittels mixed methods

Ponocny, Ivo, Weismayer, Christian; Stross, Bernadette; Dressler, Stefan
ivo.ponocny@modul.ac.at

Die vorgestellte methodische Pilotstudie zum Subjektiven Wohlbefinden (SWB) basiert auf einer simultanen Vorgabe von sowohl quantitativ als auch qualitativ orientierten Erhebungsmethoden. Im Lichte aktueller Entwicklungen zur Lebensqualitätsmessung ("Stiglitz-Report" der OECD oder die "GDP and beyond"-Initiative der EU) soll der durch mixed methods erzielbare Mehrwert debattiert werden. Insbesondere geht es um die Interpretation der quantitativen Angaben auf in der Lebensqualitätsforschung übliche Zufriedenheits- und Glücksgesamteinstufungen. Desweiteren sollte festgestellt werden, ob und wie RespondentInnen überhaupt inhaltlich klar interpretierbare Angaben über SWB treffen. Deshalb wurde – gefördert durch den Jubiläumssfond der ÖNB – eine großangelegte Studie an 10 verschiedenen Standorten in Österreich durchgeführt. Insgesamt liegen 500 qualitative Interviews zu Lebensumständen vor, über 1400 quantitative Fragebögen sowie mehr als 300 Zeitverwendungstagebücher sowie pro Standort noch je zwei Gruppendiskussionen. Die Resultate sprechen nicht nur für deutliche Zusatzinformationen durch den Einsatz von mixed methods, sondern werfen auch ein sehr kritisches Licht auf die Interpretation der Antworten auf die generellen Glücks- und Zufriedenheitsfragen. Die AutorInnen befürchten insbesondere eine Tendenz zur Schöndeutung. Die erhobenen Daten liefern aber auch eine Basis für Erklärungsmodelle, warum die quantitativen Angaben offenbar oft an manchen Problemen vorbeigehen.

Session 13 – Mixed Methods (Fr, 20.9., 15:40–17:00, HS B)

.....

Internet-basiertes Prozessmonitoring und Prozessevaluation in der Psychotherapie

Schiepek, Günter, Aichhorn Wolfgang, Strunk Guido
Guenter.schiepek@ccsys.de

In zunehmendem Maße werden in der Psychotherapie Verfahren der Internet-basierten Datenerfassung und Datenanalyse eingesetzt. Auf diesem Weg können Therapien nicht nur leichter evaluiert werden, sondern es ergeben sich auch Möglichkeiten, Therapeuten und Patienten zeitnah über die Entwicklungen zu informieren. Dargestellt werden Untersuchungen, die den therapeutischen Nutzen solcher Methoden zu belegen scheinen. Exemplarisch wird ein inzwischen insbesondere in der stationären und teilstationären Psychotherapie bewährtes System beschrieben (Synergetisches Navigationssystem), welches Selbst- und Fremdeinschätzungen in frei wählbarer Frequenz zulässt. Von mehr als 350 Patienten liegen inzwischen Daten mit täglichen Selbsteinschätzungen (Therapie-Prozess-Bogen) vor, wobei die quantitativen Daten in vielen Fällen durch Eintragungen in einem elektronischen Tagebuch ergänzt werden. Im System sind verschiedene zeitreihenanalytische Verfahren implementiert (z.B. dynamischen Komplexität, Korrelationsmatrizen, Recurrence Plots), welche über dynamische Veränderungen zeitnah Auskunft geben. Diese werden vorgestellt, einschließlich erster Befunde zu Zusammenhängen zwischen Prozessmuster und Outcome.

Session 2 – Prozessmonitoring & fMRI (Do, 19.9., 9:20–11:00, HS C)

.....

Die Kombination aus funktioneller MRT und internet-basiertem Prozessmonitoring in der Psychotherapieforschung

Schipek, Günter, Aichhorn Wolfgang, Aas Benjamin, Plöderl Martin, Kronbichler Martin, Karch Susanne

NA

Der Beitrag zeigt auf, wie funktionelle Magnetresonanztomographie in der Therapie-Prozessforschung eingesetzt werden kann. Während die Hauptanwendung von fMRT in der Klinischen Psychologie im Bereich der Evaluation und in der Identifikation neuronaler Korrelate von Störungen und/oder kognitiv-affektiven Funktionen liegt, wird hier über den wiederholten Einsatz von fMRT-Scans im Therapieverlauf berichtet. Zur Überprüfung von Hypothesen aus dem Bereich der Theorie nichtlinearer dynamischer Systeme wird tägliches Prozessmonitoring und wiederholtes fMRT kombiniert. Mit Hilfe des Internet-basierten Prozessmonitorings werden Phasen kritischer Instabilität des neuro-psychologischen Systems identifiziert, in deren Umfeld fMRT durch geführt wird. Das Stimulationsparadigma kombiniert standardisiertes mit individuellem symptomprovokativem Bildmaterial. Berichtet wird über den methodischen Ansatz eines laufenden Projekts zur stationären Psychotherapie von Patienten mit Zwangsstörungen.

Session 2 – Prozessmonitoring & fMRI (Do, 19.9., 9:20–11:00, HS C)

.....

Wirksamkeit hochschuldidaktischer Maßnahmen

Sengewald, Erik, Vetterlein, Anja
erik.sengewald@uni-jena.de

Qualitätssicherung und Qualitätsentwicklung in der Lehre haben zu Zeiten des Bologna-Prozesses und der damit verbundenen Akkreditierung der Studiengänge zunehmend an Bedeutung gewonnen. So werden zum Beispiel Kurse für Lehrende angeboten, die zur Verbesserung der Lehrqualität führen sollen. Inwiefern dieses Ziel erreicht wird betrachtet die vorliegende Studie am Beispiel der Evaluation einer hochschuldidaktischen Maßnahme. Ein mögliches Kriterium zur Beurteilung der Wirksamkeit solcher hochschuldidaktischer Maßnahmen stellt die Lehrveranstaltungsevaluation (LVE) dar. Lehrende, die an einem Kurs teilgenommen haben sollten eine höhere Lehrqualität aufweisen, als vergleichbare Lehrende, die nicht an der Maßnahme teilgenommen haben. Die Lehrqualität wird in der vorliegenden Studie mit einem Mehrebenen-Messmodell des Fragebogens zur LVE erfasst. Auf Veranstaltungsebene können 7 latente Variablen identifiziert werden, die Aspekte der Veranstaltungsqualität abbilden. Die Differenz der Mittelwerte dieser latenten Variablen zwischen den Gruppen mit und ohne hochschuldidaktische Maßnahme dient zur Evaluation der Maßnahme. Um dies zu ermöglichen wird ein entsprechendes Strukturgleichungsmodell formuliert. Ein Propensity Score Matching verbessert die Vergleichbarkeit der Lehrenden mit und ohne hochschuldidaktische Maßnahme. Im Vortrag werden sowohl die Ergebnisse der Studie vorgestellt, als auch die Methode kritisch diskutiert.

Session 1 – Evaluation I (Do, 19.9., 9:20–11:00, HS B)

.....

Auf die Plätze, fertig, los: Im Ranking gibt es nur einen Sieger und trotzdem können alle gewinnen.

Sengewald, Erik, Vetterlein, Anja
erik.sengewald@uni-jena.de

Interne Systeme zur Qualitätssicherung und Konzepte zur Qualitätsentwicklung werden an deutschen Hochschulen zunehmend etabliert. Grundlage der Qualitätssicherung und -entwicklung sind meist Evaluationen zu Beginn des Studiums, während des Studiums und nach dem Studium. Damit soll eine kontinuierliche Bewertung der Qualität eines Studienfachs ermöglicht und Veränderungen in Folge von Maßnahmen zur Qualitätsentwicklung sichtbar werden. Durch das hohe Maß an Standardisierung der Evaluationen über die verschiedenen Fächer hinweg liegt es nahe ein Ranking der verschiedenen Fächer zu erstellen. Ein Ranking auf Basis der Evaluationsergebnisse fordert jedoch, dass diese Ergebnisse ausschließlich auf die Qualität des Fachs zurückzuführen sind und somit ein fairer Vergleich zwischen den Fächern möglich ist. Im Vortrag werden die Bedingungen für Unverfälschtheit der Evaluationsergebnisse am Beispiel der Studiengangsevaluation an der FSU Jena theoretisch erläutert. Weiterhin wird gezeigt welche Probleme mit dem Ranking einhergehen und wie die Qualitätssicherung und Qualitätsentwicklung ohne Ranking erfolgreich funktionieren. Hierzu werden Evaluationsergebnisse der Fächer entsprechend aufbereitet. Mit Hilfe von latenten Klassenanalysen und multinomialen logistischen Regressionen wird eine differenzierte Deskription der Studierendenurteile ermöglicht, die der Qualitätsentwicklung dienlicher ist als ein unfairen Vergleich zu anderen Fächern.

Session 1 – Evaluation I (Do, 19.9., 9:20–11:00, HS B)

.....

Robustheit und Testgüte des *LLTM invers*

Steinfeld, Jan, Kubinger, Klaus D. ; Hohensinn, Christine
Jan.Steinfeld@univie.ac.at

Kubinger, Hohensinn, Holoher-Ertl und Heuberger (2011) schlagen z.B. im Zusammenhang mit der Bestimmung von Akzelerationseffekten in Intelligenztests vor, Fischers *LLTM invers* anzusetzen: Anstatt wie ursprünglich die Itemparameter im Rasch-Modell über entsprechende kognitive Elementaroperationen zu modellieren, werden dabei die Personenparameter hypothesenbasiert zerlegt. Mit diesem Ansatz ist eine CML-basierte Parameterschätzung möglich und Modelltests – wie der Andersen Likelihood-Ratio Test – verfügbar. Fraglich ist allerdings die Güte der Parameterschätzung sowie das asymptotische Verhalten der Verteilung der Testgröße des Likelihood-Ratio Tests: durch systematische Simulationen wird die Approximation der χ^2 -Verteilung und damit die Robustheit und Testgüte geprüft. Zusätzlich werden die Ergebnisse aus den Simulationen zur Genauigkeit der Parameterschätzung präsentiert.

Session 8 – IRT II: Modellentwicklung (Fr, 20.9., 10:20–11:00, HS C)

.....

Modellierung von Itempositionseffekten mittels generalized mixed models

Steinwascher, Merle, Meiser, Thorsten
steinwascher@uni-mannheim.de

Positionseffekte können unterschiedliche Ursachen haben, wie Ermüdungs-, Übungs- oder, im Fall von Gedächtnisspannungsaufgaben, Interferenzeffekte. Entsprechend können sich diese Effekte positiv oder negativ auf die Leistung auswirken, und sich innerhalb eines Testverlaufs verändern, z.B. positive Effekte durch Übungseffekte zu Beginn des Testverlaufs und negative Effekte durch Ermüdungs- oder Interferenzeffekte gegen Ende. Des Weiteren ist davon auszugehen, dass sich Personen bezüglich dieser Effekte unterscheiden, und dass diese Unterschiede mit der Fähigkeit einer Person in Zusammenhang stehen. Im Rahmen der vorliegenden Studie werden GMM genutzt, um komplexe Gedächtnisspannungsaufgaben hinsichtlich dieser Effekte zu untersuchen, wobei sich zeigt, dass die Annahme eines linearen Effektes eine (vereinfachende) Fehlspezifikation darstellt. Vielmehr zeigen sich ein signifikanter fester quadratischer sowie ein signifikanter fester und zufälliger kubischer Effekt. Anhand der positiven Korrelation des kubischen Effektes mit der Personenfähigkeit lassen sich Aussagen über die Verläufe von fähigen und weniger fähigen Personen treffen: Fähigere Personen weisen zu Beginn der Testung einen deutlichen Anstieg in der Leistung auf, die gegen Ende der Testung asymptotisch verläuft. Weniger fähige Personen weisen dagegen zu Beginn und gegen Ende der Testung deutliche Leistungseinbußen auf. Abschließend werden Möglichkeiten der Validierung dieser Effekte diskutiert.

Session 11 – Wachstumskurven & GMM (Fr, 20.9., 14:00–15:20, HS B)

.....

Die Anwendung von Wachstumskurvenmodellen mit latenten Klassen zur Entdeckung von Entwicklungsverläufen des auffälligen Sozialverhaltens bei Kindern und Jugendlichen

Stemmler, Mark, Lösel, Friedrich
mark.stemmler@psy.phil.uni-erlangen.de

Aus entwicklungspsychologischer Sicht ist die Erforschung früher Prozesse der Entstehung und Verfestigung von Auffälligkeiten im Sozialverhalten von großer Bedeutung. Die Erlangen-Nürnberger Entwicklungs- und Präventionsstudie (ENDPS; Lösel et al., 2009) untersucht die Entwicklung des Sozialverhaltens von Kinder und Jugendlichen mit einem bio-psycho-sozialen Modell sowie mit einem multi-method-/multi-informant-Ansatz. Beginnend im Durchschnittsalter von zirka 4 Jahren wurden Jungen und Mädchen bis zum Durchschnittsalter von zirka 13 Jahren in insgesamt sechs Erhebungswellen untersucht. Um den Verlauf sozialer Auffälligkeiten zu untersuchen, konnte auf ein Längsschnittdatensatz von N=541 Jungen und Mädchen zurückgegriffen werden. Mit Hilfe von Wachstumskurven mit latenten Klassen sogenannte General Growth Mixture Models (GGMM; Reinecke, 2006; McArdle, 2008) werden die Angaben von Erzieherinnen, Lehrerinnen und Eltern zum Sozialverhalten von Jungen und Mädchen analysiert. Im Vortrag werden der Ansatz der GGMM erklärt sowie deren Vorzüge und Probleme diskutiert.

Session 11 – Wachstumskurven & GMM (Fr, 20.9., 14:00–15:20, HS B)

.....

Risiken und Nebenwirkungen optimal-selektierter Statistiken am Beispiel der modellbasierten rekursiven Partitionierung von Rasch- und Partial-Credit-Modellen

Strobl, Carolin

carolin.strobl@uzh.ch

Die modellbasierte rekursive Partitionierung kann zum Aufdecken von Parameter-Invarianzen in psychometrischen Modellen verwendet werden, wie z.B. zur Diagnose von Differential-Item- und Differential-Step-Functioning im Rasch- und Partial-Credit-Modell. Dabei werden Personen-Kovariablen (z.B. Alter und Geschlecht) eingesetzt, um Gruppen von Personen mit unterschiedlichen Aufgaben-Parametern zu identifizieren. Ein Vorteil dieses Verfahrens gegenüber herkömmlichen Modelltests für vorgegebene Gruppen (die zur Aufteilung stetiger Variablen üblicherweise den Median als Bruchpunkt verwenden) ist, dass auch Gruppen entdeckt werden können, die sich aus anderen Bruchpunkten oder mehreren Kovariablen ergeben (z.B. Männer über 40 Jahre vs. dem Rest der Bevölkerung). Die zentrale methodische Herausforderung bei der rekursiven Partitionierung stellt dabei die Berücksichtigung der optimalen Auswahl des Bruchpunktes bei Kovariablen mit mehr als zwei Ausprägungen dar: Durch das Ausprobieren aller möglichen Bruchpunkte würde eine Testentscheidung anhand der herkömmlichen statistischen Verteilung zu einer Inflation des alpha-Niveaus führen. Dies muss bei der Konstruktion des Algorithmus z.B. durch die Herleitung der (asymptotischen oder exakten) Verteilung der optimal selektierten Statistik berücksichtigt werden. Der Vortrag zeigt die Problematik der optimal-selektierten Statistiken am Beispiel der modellbasierten rekursiven Partitionierung sowie am einfacheren Beispiel des χ^2 -Tests auf. Aus diesen Beispielen wird klar, dass es sich hierbei um ein grundlegendes statistisches Problem handelt, das nicht nur in der rekursiven Partitionierung auftritt - und deshalb in der computationalen Statistik und Biometrie einen zentralen Forschungsgegenstand darstellt.

Session 14 – IRT V: Modelltests & Parameterschätzung (Fr, 20.9., 15:40–17:00, HS C)

.....

Bessere Lehre durch bessere Ergebnisdarstellung der Lehrevaluation?

Vetterlein, Anja, Sengewald, Erik

anja.vetterlein@uni-jena.de

Nach Marsh (2007) ist eine Funktion der Lehrevaluation das (diagnostische) Feedback und soll der Verbesserung der Lehre dienen. Dabei reicht der bloße Empfang nicht aus, sondern die (statistischen) Ergebnisse müssen auch verstanden werden, um die richtigen Verbesserungsmaßnahmen abzuleiten (Müller, 2010). Mit dem Ziel, die Ergebnisse aus der Lehrevaluation wirksamer abzubilden, wurde eine neue Darstellungsform für die Ergebnisberichte entwickelt. Im Rahmen einer Treatment-Studie erfolgt die randomisierte Zuweisung der alten vs. der neuen Darstellungsform zu den Dozenten. Der empirische Vergleich der zwei Darstellungsformen, soll die Frage beantworten, unter welcher Treatment-Bedingung sich die Lehre mehr verbessert. Als abhängige Variable dient das Evaluationsergebnis der jeweils nachfolgenden Veranstaltung jedes Dozenten. Zum Test auf bedingte Effekte werden weitere Einflussvariablen berücksichtigt (u.a. zurückliegende Lehrveranstaltungsevaluation als Pretest des Dozenten, Pretest der Erfahrung mit Evaluationsberichten, Fakultätszugehörigkeit, Geschlecht).

Session 1 – Evaluation I (Do, 19.9., 9:20–11:00, HS B)

.....

Methodisches Vorgehen in der Evaluation der Nachhaltigkeit Lokaler Aktionspläne gegen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus in Deutschland

Volf, Irina

irina.volf@iss-ffm.de

Im Jahre 2007 hat das deutsche Bundesministerium für Familie, Senioren, Frauen und Jugend das Programm VIELFALT TUT GUT. Jugend für Vielfalt, Toleranz und Demokratie aufgelegt und damit 90 Lokale Aktionspläne bundesweit ins Leben gerufen. Ihre Aufgabe ist es, auf der Basis von Analysen lokaler Problemlagen integrierte lokale Handlungsstrategien zu erarbeiten, die auf eine gemeinwesenorientierte Auseinandersetzung mit den Themen Rechtsextremismus, Fremdenfeindlichkeit und Antisemitismus in den unterschiedlichen Handlungsebenen abzielen. 2010 bekamen 84 dieser Aktionspläne eine Anschlussfinanzierung im Rahmen des Nachfolgeprogramms Toleranz Fördern – Kompetenz Stärken, um die Verstetigung der integrierten lokalen Strategien zu sichern. Mit der wissenschaftlichen Begleitung und der Evaluation der Aktionspläne wurden das Institut für Sozialarbeit und Sozialpädagogik e.V. und Camino gGmbH beauftragt. Ein mehrdimensionales Nachhaltigkeitskonzept wurde in der ersten Förderphase qualitativ entwickelt und quantitativ getestet. In der zweiten Förderphase wurde die Evaluation als Längsschnittanalyse angelegt, durch Indexbildung fortentwickelt und mit der Latent Class Analysis ausgewertet. Das zirkuläre Zusammenführen der qualitativen und quantitativen Befunde ermöglichte es, Ergebnisse mit hoher Qualität auf einem zunehmenden Komplexitätsniveau darzustellen. Der Vortrag fokussiert auf die methodologische Vorgehensweise der Untersuchung und stellt die ersten Befunde vor.

Session 13 – Mixed Methods (Fr, 20.9., 15:40–17:00, HS B)

.....

Strukturgleichungen in Onyx

von Oertzen, Timo, Brandmaier, Andreas; Tsang, Siny

timo@virginia.edu

Onyx ist eine neues offen zugängliches SEM-Programm zum Erstellen von Strukturgleichungsmodellen als Pfaddiagramm und Schätzen von Parametern. Onyx bietet eine intuitive graphischen Benutzeroberfläche, die sowohl für Anfänger im Bereich von Strukturgleichungsmodellen als auch für die Lehre flexibel einsetzbar ist. In Onyx erstellte Pfaddiagramme können unmittelbar in Skriptsprachen wie OpenMx, lavaan, oder Mplus angezeigt und exportiert werden. Zusätzlich zu den graphischen Eigenschaften verfügt Onyx über effiziente Parameterschätzung, die es ermöglicht, ohne Startwerte vom Benutzer verschiedene lokale Minima zu finden und miteinander zu vergleichen. Onyx erlaubt das Fitten von Multi-Gruppen-Modellen, die Verwendung von definition variables, und Full Information Maximum Likelihood beim Parameterschätzen mit fehlenden Variablen. Im diesem Vortrag werden wir die zugrundeliegende Theorie von Onyx erläutern und das Programm demonstrieren.

Session 3 – SEM I/Pfadmodelle: Anwendung & Caveats (Do, 19.9., 13:20–15:00, HS B)

.....

Diffusionsmodellanalysen mit fast-dm

Voss, Andreas, Lerche, Veronika

andreas.voss@psychologie.uni-heidelberg.de

Der Einsatz stochastischer Diffusionsmodelle zur Analyse von Reaktionszeitdaten hat in den letzten Jahren besonders im Bereich der kognitiven Psychologie stark zugenommen. Aber auch in anderen Disziplinen, wie etwa der Sozialpsychologie, der klinischen Psychologie oder der diagnostischen Psychologie finden Diffusionsmodelle erste Anwendungen. Die Diffusionsmodellanalyse erlaubt es aus den Reaktionszeitverteilungen von korrekten Antworten und Fehlern eine Reihe von Parametern zu schätzen, die spezifische kognitive Komponenten der Informationsverarbeitung abbilden. Von besonderem Interesse sind dabei die Maße für (a) die Menge der berücksichtigten Information (Schwellenabstand), (b) die Geschwindigkeit der Informationsaufnahme (Drift) und (c) die Dauer nicht-entscheidungsbezogener Prozesse. Im vorliegenden Vortrag wird eine Weiterentwicklung des Auswertungsprogramms fast-dm zur Schätzung der Diffusionsmodellparameter vorgestellt. Insbesondere wird dabei die Robustheit und Effizienz der Parametersuche bei Verwendung unterschiedlicher Optimierungskriterien (Maximum Likelihood, χ^2 , Kolmogorov-Smirnov) verglichen.

Session 5 – Multinomiale und Diffusionsmodelle (Do, 19.9., 15:20–17:00, HS B)

.....

RWiener package

Wabersich, Dominik, Vandekerckhove, Joachim

dominik.wabersich@gmail.com

The RWiener package provides functions to implement the Wiener diffusion process model in R. The functions include the density function, the cumulative distribution function, a quantile function, and a function to do random sampling. The computations are performed highly efficient algorithms taken from recent scientific papers. RWiener also provides additional functions for further inference or to do parameter estimation.

Session 5 – Multinomiale und Diffusionsmodelle (Do, 19.9., 15:20–17:00, HS B)

.....

Nichtlineare Effekte in Multimethoden-Analysen

Werner, Christina S., Zapf, Dieter; Schermelleh-Engel, Karin; Strobl, Carolin

c.werner@psychologie.uzh.ch

Beim Analysieren von Multimethoden-Daten stellt sich die Aufgabe, methodenbedingte Kovarianz zwischen Messungen verschiedener Konstrukte angemessen zu berücksichtigen. Kritisch ist dies besonders bei Analysen nichtlinearer Effekte: Interaktions- und quadratische Terme können untereinander stark zusammenhängen, so dass bei messfehlerbehafteten Variablen methodenbedingte Kovarianz zu zusätzlichen Verzerrungen dieser Effekte führt. Dagegen erlauben nichtlineare Multimethoden-Strukturgleichungsmodelle (SEM) unverzerrte Schätzungen nichtlinearer Effekte auf der Ebene latenter Variablen. Simulationen zeigen, dass selbst unter suboptimalen Bedingungen wie schief verteilten Variablen Strukturgleichungsmodelle einer vermeintlich voraussetzungsärmeren Analyse beobachteter Variablen überlegen sind. Nichtlineare Multimethoden-SEM erlauben somit valide Analysen für Forschungsfelder, in denen die Kombination verschiedener Methoden wünschenswert ist, wie an einem Beispiel zu Arbeitsstress demonstriert wird.

Session 7 – SEM II: Nichtlineare Effekte (Fr, 20.9., 10:00–11:00, HS B)

.....

Eine alternative Methode zur Modellierung der Fähigkeitsverteilung in der Population bei Large-scale Educational Surveys

Wetzel, Eunike, Xu, Xueli; von Davier, Matthias
eunike.wetzel@uni-bamberg.de

In Large-scale Educational Surveys wird operational ein latentes Regressionsmodell verwendet, um den Mangel an kognitiven Informationen zu kompensieren. Das latente Regressionsmodell basiert dabei auf Hauptkomponenten, die aus Hintergrunddaten extrahiert werden. Dieses Modell hat wichtige Nachteile, z.B. beim Umgang mit fehlenden Werten. In diesem Beitrag wird eine alternative Methode vorgestellt, die mit einer Latenten Klassenanalyse (LCA) Gruppen identifiziert, mithilfe derer Schüler hinsichtlich ihrer Hintergrundcharakteristika klassifiziert werden können. Diese Klassifikationen werden als Prädiktoren in das latente Regressionsmodell eingeführt. Ziel dieser Studie war es zu untersuchen, ob diese Methode ähnliche Ergebnisse liefert wie die operationale Prozedur. Verschiedene LCA-basierte Ansätze unterschieden sich danach, ob sie manifeste Klassenzugehörigkeiten oder Wahrscheinlichkeiten der Zugehörigkeit zu den latenten Klassen als Prädiktoren verwendeten. Insgesamt geben die LCA-Ansätze die Gruppenmittelwerte und Standardabweichungen der operationalen Prozedur gut wider. Weiterhin korrelieren die Mittelwerte und Standardabweichungen der a-posteriori-Verteilungen, die zur Ziehung von Plausible Values verwendet werden, hoch zwischen der operationalen Prozedur und den LCA-Ansätzen. Im latenten Regressionsmodell Prädiktoren zu verwenden, die auf einer LCA der Hintergrunddaten basieren, scheint daher eine brauchbare Alternative zum operationalen Hauptkomponentenansatz zu sein.

Session 8 – IRT II: Modellentwicklung (Fr, 20.9., 10:20–11:00, HS C)

.....

Resampling-Tests zur Kausalitätsprüfung in linearen Regressionsmodellen

Wiedermann, Wolfgang, Hagmann, Michael; Kossmeyer, Michael; von Eye, Alexander
wolfgang.wiedermann@univie.ac.at

Dodge und Rousson (2001) diskutieren asymmetrische Eigenschaften des Pearson-Korrelationskoeffizienten und zeigen, dass die höheren Momente der abhängigen (Y) und unabhängigen Variablen (X) zur Bestimmung der directional dependence im Kontext linearer Regressionsmodelle herangezogen werden können. Die vorliegende Arbeit erweitert dieses Konzept und diskutiert Verteilungseigenschaften der Residualterme zweier konkurrierender Modelle (X prädiziert Y vs. Y prädiziert X). Ausgehend von den dritten Momenten der Residuen werden drei Resampling-Tests zur inferenzstatistischen Modellselektion vorgestellt. Die Leistungsfähigkeit der Verfahren wurde mittels Simulationsexperimenten überprüft, wobei die folgenden Simulationsfaktoren systematisch variiert und gekreuzt wurden: Schiefe der unabhängigen Variable, Grad der Korrelation zwischen X und Y, Stichprobengröße und Verteilung der Residuen (Normal- vs. Laplace-Verteilung). Mit steigender Stichprobengröße und steigender Schiefe der UV steigt generell auch die Power der Tests jenes Modell zu identifizieren, welches dem wahren datengenerierenden Prozess entspricht. Mit steigender Korrelation sinkt die Selektionsleistung der Verfahren. Diese Effekte gelten für normalverteilte sowie für Laplaceverteilte Residuen. Die Anwendung der Tests wird anhand eines realen Datenbeispiels demonstriert. Die Autoren stellen eine Implementierung der Resampling-Tests in R bereit. Implikationen für die Praxis sowie kritische Aspekte werden diskutiert.

Session 4 – Regression – FA – VA – CA (Do, 19.9., 13:20–15:00, HS C)

.....

Profilanalyse und Person-Fit in der Einzelfalldiagnostik: Ein Anwendungsbeispiel aus dem berufspsychologischen Service der Bundesagentur für Arbeit.

Yousfi, Safir

safir.yousfi@arbeitsagentur.de

Die Analyse von Unterschieden zwischen den Normwerten zu verschiedenen Merkmalsbereichen (Testprofil) spielt eine wichtige in der Einzelfalldiagnostik. Dabei sollte allerdings sichergestellt werden, dass Unterschiede, die so klein sind, dass sie sich auch durch Messfehler erklären lassen, nicht inhaltlich interpretiert werden. Andererseits können extreme Differenzen zwischen den Normwerten verschiedener Tests Zweifel an der Validität der Testwerte begründen. Extreme Profile von Normwerten können auch als Indikator für einen mangelnden Personen-Fit von solchen Messmodellen interpretiert werden, die die Bildung eines aggregierten Testwerts rechtfertigen. Um die angemessene Interpretation von Normwertdifferenzen zu unterstützen, wurden in das computergestützte Testsystem des berufspsychologischen Service der Bundesagentur für Arbeit Elemente der Profilanalyse und Person-Fit Maße integriert. Die entsprechenden statistischen Tests untersuchen das Muster der Normwerte darauf, ob Unterschiede zwischen den Normwerten (a) durch Messfehler erklärt werden können und (b) ob diese Unterschiede so groß sind, dass sie Zweifel an der Validität der Testwerte begründen. Die Methoden zur zufallskritischen Absicherung des Testprofils beruhen auf einer multivariaten Verallgemeinerung des Konzepts der kritischen Differenz. Außergewöhnliche Normwertprofile werden mit einer multivariaten Maß für diagnostische Valenz und einem nachgeschalteten bayesianischen Ausreißertest identifiziert.

Session 15 – Evaluation II (Sa, 21.9., 9:30–10:50, HS B)

.....

Auf der Suche nach Zusammenhängen zwischen Sicherheitskultur und Sicherheitsempfinden der Patienten in Krankenhäusern - Datamining in der Praxis des Gesundheitswesens

Zinn, Winfried

zinn@metrik.de

Das Thema Risikomanagement gewinnt auch im deutschsprachigen Raum deutlich an Bedeutung (Koppenberg, Gausmann, & Henninger, 2012). Ein Aspekt ist die gelebte Sicherheitskultur (SK) der Mitarbeiter. Andererseits rückt die Patientensicht immer stärker in den Vordergrund (Zinn, 2010b) S.287. Für die Geschäftsführung der Krankenhäuser stellt sich die Frage, ob es einen Einfluss der SK auf das Sicherheitsempfinden der Patienten gibt. In 7 Kliniken wurde zeitgleich die SK und das Patientensicherheitsempfinden gemessen. Die Daten wurden anschließend auf Stationsebene aggregiert. 851 Mitarbeiter füllten eine Sicherheitskulturfragebogen aus. 5.081 Patientenfragebögen zum Risiko und Sicherheitsempfinden (Gausmann, Göpfert, & Zinn, 2008) während des Krankenhausaufenthaltes waren auswertbar. Bei 73 Stationen lagen sowohl Fragebögen zum Sicherheitsempfinden der Patienten, als auch zur SK der Station vor. Eine erste Korrelationsanalyse ergab 41 höchstsignifikante Ergebnisse von den 14.148 zu interpretierenden Korrelationen. Für das Datamining wurden der Einfluss von 324 potentielle Variablen des Sicherheitskulturfragebogen auf 41 Variablen des Patientensicherheitsfragebogen mittels random forrest und boosting geprüft. Die nach der Inhaltsanalyse durchgeführte Regressionsanalyse ergab, dass die SK keinen direkten Einfluss auf das Sicherheitsempfinden hat. Erfolgreiches Risikomanagement bedarf somit den Blick auf beide Aspekte.

Session 15 – Evaluation II (Sa, 21.9., 9:30–10:50, HS B)

.....

Analyse von Gruppeneffekten in Mehrebenenmodellen mit Hilfe von MCMC-Verfahren: Befunde einer Simulationsstudie zur Evaluation verschiedener Prior-Verteilungen

Zitzmann, Steffen, Lüdtke, Oliver, Robitzsch, Alexander
zitzmans@hu-berlin.de

In Mehrebenenanalysen werden zur Untersuchung von Gruppeneffekten häufig Gruppenmerkmale durch die Aggregation von Individualmerkmalen erfasst (z.B. mittlere Schulleistung). Dabei stellt der beobachtete Gruppenmittelwert in bestimmten Datenkonstellationen (z.B. geringe Anzahl von Beobachtungen innerhalb der Gruppe) eine unreliable Messung des wahren Gruppenmittelwertes dar. Es wurde ein Mehrebenenmodell (latent covariate model; Lüdtke et al., 2008) vorgeschlagen, in dem der wahre Gruppenmittelwert als latente Variable konzeptualisiert wird, um für die Unreliabilität zu korrigieren. In Simulationsstudien zeigte sich, dass der Schätzer des Kontexteffektes in problematischen Datenkonstellationen (geringe ICC sowie geringe Anzahl von Ebene-2 Einheiten) variabler war als der des klassischen Ansatzes, der den beobachteten Mittelwert verwendet. In dieser Studie wird eine Bayesianische Variante des Modells mit Hilfe einer Simulationsstudie evaluiert. Im Fokus stehen verschiedene Prior-Verteilungen für die Varianz bzw. Standardabweichung der Kovariaten zwischen den Gruppen. Es zeigte sich, dass das Modell mit einer leicht informativen inversen Gamma-Prior hinsichtlich des RMSE am besten abschnitt. Mit zunehmender Information für den Gruppenmittelwert der Kovariaten (steigende ICC oder Gruppengröße) verringerte sich der Einfluss der Wahl der Prior auf den Schätzer des Gruppeneffektes und die Ergebnisse näherten sich dem Maximum-Likelihood-Schätzer. Anhand eines realen Datenbeispiels wird illustriert, wie das Modell in dem Programm WinBUGS geschätzt werden kann.

Session 9 – Mehrebenenmodelle (Fr, 20.9., 11:20–12:40, HS B)

.....

Index

Alexandrowicz , 3
Baghaei , 3
Bollmann , 4
Born , 4
Brandt , 5
Buchholz , 5
Burger , 6
Carstensen , 6
Doebler , 7
Fartacek , 7
Fenzl , 8
Fischer , 8
Fondel , 9
Forthmann , 9
Frey , 10
Fuchs , 10
Gula , 11
Hagemann , 11
Hartig , 12
Hecht , 12
Heene , 13
Heine , 13
Hohensinn , 14
Köhler , 16
Keller , 14
Kerwer , 15
Khorrarnadel , 15
Klein , 16
Koller , 17
Krenn , 17
Kutscher , 18
Leuschner , 18
Lubbe , 19
Maier , 19
Martin , 19
Mayerl , 20
Meiser , 20
Morawetz , 21
Mutz , 21
Niemeier , 22
Nussbeck , 22
Pfaffel , 23
Pfeiffer , 23
Plieninger , 24
Pohl , 24
Ponocny , 25
Schipek , 25, 26
Sengewald , 26, 27
Steinfeld , 27
Steinwascher , 28
Stemmler , 28
Strobl , 29
Vetterlein , 29
Volf , 30
von Oertzen , 30
Voss , 31
Wabersich , 31
Werner , 31
Wetzel , 32
Wiedermann , 32
Yousfi , 33
Zinn , 33
Zitzmann , 34