Claus H. Carstensen und Jürgen Rost

IPN Kiel

MULTIRA

ein Programmsystem zur Analyse mehrdimensionaler Rasch-Modelle

Handbuch zum Computerprogramm **MULTIRA**

Einführung – Theorie & Praxis – Referenz

Zur Programmversion v1.45,

Kiel, Dezember 2000

Institut für die Pädagogik der Naturwissenschaften

Olshausenstrasse 62

24116 Kiel

Inhalt

Vorweg

Zur Schreibweise MultiRa und MULTIRA

Zuerst eine Bemerkung zur mehrfachen Verwendung des Namens Multira: Zum einen bezeichnet MultiRa ein psychologisches Testmodell und den von uns dazu entworfenen und in diesem Programm umgesetzten Algorithmus, zum anderen ist MULTIRA gleichzeitig der Name dieses Computerprogrammes, welches noch weitere Algorithmen und Rechenprogramme anbietet. Das Computerprogramm wird im folgenden durch Großschreibung MULTIRA gekennzeichnet, das Testmodell und der Algorithmus durch die Schreibweise MultiRa.

Über dieses Handbuch

Dieses Handbuch soll in die Arbeit mit dem Programm MULTIRA einführen und die Arbeit mit dm Programm begleiten und unterstützenn. Dazu sind relativ viele Kapitel möglichst praxisnah geschrieben und theoretische Erläuterungen fallen eher kurz aus. Für eine detaillierte Einführung in theoretische Hintergründe werden folgende Publikationen empfohlen:

- Carstensen, Claus H., (2000). *Mehrdimensionale Testmodelle mit Anwendungen aus der pädagogisch-psychologischen Diagnostik*. Kiel: IPN, Schriftenreihe 171.
- Rost, J. & Carstensen, C. H., (in press). Multidimensional Rasch Measurement via Item ComponentModels and Faceted Designs. Applied Psychological Measurement.

Dieses Handbuch gliedert sich in sechs Kapitel, die sich neben inhaltlichem auch darin unterscheiden, auf welche Weise sie dem Leser nützlich sein mögen. Das **erste Kapitel** (S. 6) gibt einen schnellen Einstieg in die Arbeit mit dem MultiRa-Algorithmus. Anhand eines Beispieles werden einige wesentliche Elemente der Programmbedienung und der Ergebnissausgabe vorgeführt.

Die folgenden Kapitel sind ausführlicher als das erste gehalten, sie sollen dem Leser zum Einarbeiten in das MultiRa-Modell und in das Programm ebenso wie zum Nachschlagen dienen. **Kapitel 2** (S.11) beschreibt das MultiRa-Modell und die Schätzalgorithmen theoretisch; weitere im Programm MULTIRA implementierte Verfahren werden in **Kapitel 3** (S.22) dargestellt.

Kapitel 4 (S. 24) beschreibt den Umgang mit MULTIRA's Benutzeroberfläche und erläutert kurz die zur Verfügung stehenden Optionen. Das **fünfte Kapitel** (S.38) erläutert eine Ergebnisdatei; es handelt sich dabei um die vollständige Ergebnisdatei zu dem Beispiel aus Kapitel 1. Das **sechste Kapitel** (S.45) erläutert in weiteren Beispielen verschiedene Anwendungen des MultiRa-Modells, um so das Verständnis in die Arbeit mit dem Programm vertiefen.

Dieses Handbuch liegt in zwei Versionen vor, in einer gedruckten Fassung und als Onlinehandbuch im Adobe Acrobat PDF-Format. Benutzer des Onlinehandbuches können die Lesezeichen im PDF-Dokument verwenden, um durch das Handbuch zu „navigieren“. Alle Querverweise (hauptsächlich auf Seitenzahlen) stehen im Onlinehandbuch als Hyperlinks zur Verfügung. Ein Stichwortindex (S. 48) soll das Suchen nach bestimmten Begriffen erleichtern.

Installation und Systemanforderungen

Das Setupprogramm wird durch den Aufruf der Datei >multira_setup_de.exe< bzw. der Datei >multira_setup_eng< gestartet. Folgen Sie den Dialogen des Setupprogrammes. Sie werden gebeten, ein Verzeichnis auf der Festplatte für die Programmdateien anzugeben und ein Programmstartgruppe für die Verknüpfungen auszuwählen.

MULTIRA läuft unter Windows 95b oder neuer. Es belegt etwa 3 Mb Speicher auf der Festplatte, auf dem Rechner sollten mind. 8 Mb Arbeitsspeicher vorhanden sein.

Mit der Installation werden Programm- und Beispieldateien in das auswählbare MULTIRA Programmverzeichnis kopiert und die Dateien ftn90.dll, salflibc.dll, das elektronische Handbuch und eine Verknüpfung mit der MULTIRA-Homepage ins Windows-Verzeichnis kopiert. Werden Dateien mit diesen Namen ebenfalls durch andere Programme genutzt, kann es zu Fehlern kommen (wurde selten unter NT4 mit der Datei ftn90.dll beobachtet)

1. Eine kurze Einführung

1.1 Die Rechenprogramme in MULTIRA

Nebem dem MultiRa-Algorithmus sind weitere Algorithmen und Rechenprogramme in MULTIRA enthalten. Dazu gehört ein Simulationsprogramm zur Erzeugung von Datensätzen nach dem MultiRa-Modell (S.36), der Algorithmus MKAT zur Schätzung der Parameter des mehrkategoriellen Rasch-Modells (S.22) und ein Algorithmus zum dichotomen LLTM (S.23). Der generalisierte Martin-Löf Test kann mit dem CML Multira-Algorithmus (S.22) durchgeführt werden.

1.2 Eine Beispielanwendung des MultiRa-Algorithmus

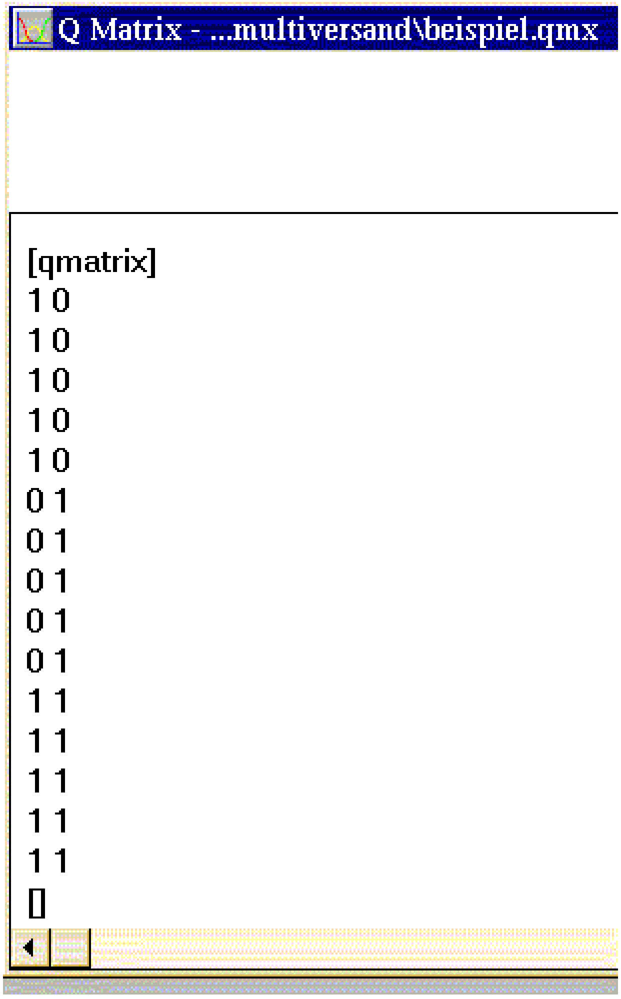
In diesem Kapitel wird an einem Beispiel vorgeführt, wie mit dem Programm MULTIRA ein Datensatz analysiert werden kann. Dazu wird das Datenbeispiel (>beispiel.*<) benutzt, welches mit den Installationsdateien mitgeliefert wird und ins Programmverzeichnis kopiert wurde.

Es handelt sich bei diesem Datenbeispiel um simulierte Daten für einen hypothetischen Test, der zwei unterschiedliche Leistungen mißt. Dabei gibt es Items, die jeweils allein mit einer Leistungsfähigkeit gelöst werden können und auch Items, die beide Leistungen erfordern. Dies könnten zum Beispiel zwei unterschiedliche kognitive Strategien sein oder zwei Fähigkeiten, die sich wie "reasonning" und "space" in einigen Intelligenzaufgaben überschneiden.

Der hypothetische Test umfasst fünfzehn Items, in der Datei >beispiel.qmx< ist die Designmatrix Q gespeichert, welche die mehrdimensionale Struktur der Items wiedergibt. In der Datei >beispiel.dat< sind simulierte Antworten für 500 Personen abgespeichert.

Um eine Datenanalyse rechnen zu können, müssen beide Dateien geöffnet werden. Mit dem Befehl 

öffnet man die Designmatrix; der folgende Bildschirmausschnitt zeigt die Designmatrix in einem Q-Matrix-Fenster von MULTIRA geöffnet.



Die ersten fünf Items sprechen die erste Leistungsfähigkeit an, die Items sechs bis zehn die zweite und die Items elf bis fünfzehn die dritte. Nach dieser Designmatrix wurde das Datenbeispiel generiert.

Mit dem Befehl 

wird der Datensatz geöffnet. Nach Auswahl einer Datei aus einem Standard-Dialog-Fenster erscheint das Dialog-Fenster "Kennwerte der Daten", welches einige Informationen über den Datensatz erfragt.

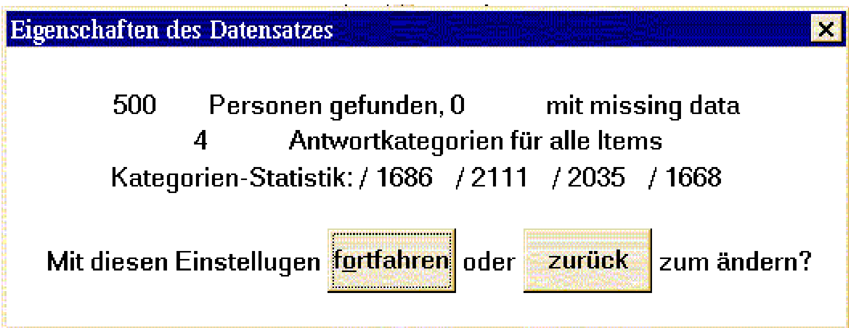
Zuerst muß die Zahl der Items angegeben werden und die Zahl der Zeichen, die je Item abgspeichert wurde; für das Programm MULTIRA müssen alle Items in derselben Spaltenbreite abgespeichert sein. Im Datenbeispiel sind die Items zweispaltig abgespeichert.

Sind die Antworten, wie im Falle des Beispieldatensatzes, mit Zahlen kodiert, kann im Kasten "Antwortformat"die entsprechende Option "ordinal kodierte Itemantworten" gewählt werden. Dabei müssen die niedrigste und die höchste auftretende Kategorie angegeben werden. Antworten ausserhalb dieses Intervalls werden als fehlende Werte eingelesen. Da die Antwortkategorien so rekodiert werden, daß die niedrigste Kategorie auf Null fällt, ist es wichtig, genau die niedrigste Kategorie anzugeben. Die Angabe der höchsten Kategorie muß lediglich höher sein, als die der auftretenden gültigen Antworten.

Fehlende Werte gibt es im Beispiel nicht, daher ist die Einstellung dieser Option unwesentlich. Sind alle Angaben getroffen, wird der Datensatz mit der Schaltfläche 

auf die Anzahl der Personen und fehlende Werte hin untersucht.

Das Ergebnis wird in einem Fenster angezeigt, sollten die Angaben nicht korrekt sein, kann man zum vorherigen Fenster zurückgehen.



Nach Betätigen der "Fortfahren"-Schaltfläche wird der Datensatz gemäß den Einstellungen eingelesen und rekodiert gespeichert. Damit steht er für Analysen zur Verfügung.

Mit dem Befehl 

öffnet man ein Dialog-Fenster mit mehreren Registerkarten, in denen man Optionen wählen kann; für das Datenbeispiel können die Voreinstellungen benutzt werden.

Nach Betätigen der Schaltfläche 

rechts im Fenster beginnt MULTIRA zu rechnen. Während der Parameterschätzungen informiert ein Fenster über die Fortschritt der Iterationen. Neben anderen wird die Differenz der Loglikelihood der letzten beiden Iterationen angezeigt, wird diese kleiner als der vorher in der Registerkarte "Algorithmus" eingestellte Wert oder wird die dort eingestellte maximale Zahl von Iterationen überschritten, beendet das Programm die Parameterschätzung und berechnet verschiedene Statistiken für die Ergebnisausgabe.

Mit dem Befehl 

kann die Ergebnisdatei in einem Fenster geöffnet werden, das Fenster wird maximiert dargestellt. Zuerst werden in der Ergebnisdatei einige Eigenschaften des Datensatzes wiedergegeben und die Designmatrix dargestellt.

Etwas weiter werden Parameterzahlen dargestellt, die folgenden Zeilen geben an, daß 20 Personen in einer oder beiden Dimensionen einen Maximalscore oder einen Minimalscore aufweisen und daher für die unbedingte Maximum Likelihood-Schätzung ausgenommen werden mußten.

```
Anzahl der Personen im Datensatz                     :   500
... mit mindestens einem Minimal- oder Maximalscore :    20

... in der folgenden Parameterschätzung              :   480
```

Zur Schätzung der weighted Maximum Likelihood Schätzer für die Personenparameter werden diese Personen wieder einbezogen.

Nach weiteren Parameterzahlen und den Kategorienhäufigkeiten werden die Iterationen protokolliert, dies kann mit der Option "Kurze Ergebnisdarstellung" in der Registerkarte "Ergebnisse" abgeschaltet werden.

Die Personenparameter werden in einer separaten Datei gespeichert, es folgen einige deskriptive Kennwerte der Parameterschätzungen, wie Mittelwert und Streuung der Personenparameter und die Reliabilität der Dimensionen, die sich anhand der Itemparameter-Schätzfehler berechnet.

```
### Mittelwerte, Streuungen & Reliabilität für die Personenparameter ###
Komp.       M       SD      Rel    Mittelwert der mittleren Itemparameter in j
   1    0.092    1.608    0.878    0.000
   2   -0.273    1.623    0.886    0.000
```

Die Itemparameter werden in einer Tabelle dargestellt, zunächst die Schwellenparameter und dann nach dem M: "Itemmitteparameter", die sich aus der Summe der Schwellenparameter dividiert durch deren Anzahl ergeben.

In der folgenden Tabelle ist die unbedingte Loglikelihood angegeben und der Wert, der sich für die unbedingte Loglikelihood ergibt, wenn die weighted ML-Schätzer zugrungegelegt werden. Das geometrische Mittel gibt die Durchschnittliche Wahrscheinlichkeit einer Antwort einer Person auf ein Item an. Die Verwendung der unbedingten Likelihood zur Modellgeltungskontrolle ist problematisch, da sie viele inzidentelle Parameter enthält. Besser geeignet ist das folgend beschriebene Bootstrap-Verfahren.

Im letzten Abschnitt werden zwei χ^2-Statistiken ausgegeben, die den Unterschied zwischen den nach den Modellparametern zu erwartenden Patternhäufigkeiten und den tatsächlich beobachteten Patternhäufigkeiten zur Grundlage haben. Bei der nach der Klassifikation von Cressie und Read (1982, siehe auch von Davier, 1997) bezeichneten Statistik CR(1) handelt es sich um den Pearson χ^2-Wert, CR(2/3) ist eine Statistik von Cressie und Read.

```
### Abweichung zw. erwarteten und beobachteten Patternhäufigkeiten ###
        df           CR(1)      CR(2/3)
  1073740820    852.737E+06   5.457E+06
```

Da im Datenbeispiel wesentlich mehr Pattern untersucht werden, als Antwortmuster (=Personenzahl) beobachtet wurden, kann die asymptotische χ^2-Verteilung dieser Statistiken nicht angenommen werden. Es sollte das Bootstrap-Verfahren benutzt werden, welches bei gegebenen Parametern Datensätze resimuliert, deren Parameter schätzt, die beiden χ^2-Statistiken berechnet und auf diese Weise empirisch eine Prüfverteilung erstellt.

Soweit diese Kurzbeschreibung der Analyse des Datenbeispieles; zum MultiRa-Modell steht Eine ausführlichere Darstellung des Modells gibt Kapitel 2 (S.11) und die Bedienung des Programms erläutert Kapitel 4 (S. 24). Eine ausführlichere Beschreibung der Ergebnisdatei findet sich in Kapitel 5 (S.38) und weitere Beispiele werden im nächsten Abschnitt und in Kapitel 6 (S.45) erläutert.

1.3 Ein Beispiel mit CML Schätzung

2. Das *Multi*dimensionale Itemkomponenten *Ra*sch Modell *MultiRa*

Im folgenden wird das Testmodell, welches im Programm MULTIRA seine Anwendung findet, dargestellt und es werden theoretische Grundlagen weiterer Programmbestandteile beschrieben.

2.1 Einleitung

Das MultiRa-Modell formalisiert die Annahme, daß mehrere Personeneigenschaften auf additive Weise gleichzeitig am Zustandekommen einer Antwort auf ein Item beteiligt sind. Dieses Modell leitet sich aus dem Rasch Modell ab: die Rasch-Meßtheorie erhält damit ihre mehrdimensionale Erweiterung, ähnlich wie die klassische Testtheorie durch die Faktorenanalyse ihre mehrdimensionale Erweiterung erhalten hat.

2.2 Darstellung des Modells

Zunächst wird das MultiRa-Modell vorgestellt und dann kurz vorgeführt, wie es sich aus dem einfachen logistischen Modell von Rasch (1960) herleiten läßt. Doch erst einmal zur Modellgleichung des MultiRa -Modells, Gleichung 1:

Gleichung 1
$$p(X = x|v,i) = \frac{\exp\left(x\sum_{j=1}^{J} q_{ij}\theta_{vj} - \sigma_{ix}\right)}{\sum_{s=0}^{m}\exp\left(s\sum_{j=1}^{J} q_{ij}\theta_{vj} - \sigma_{is}\right)},$$
MULTIRA-Modell

Dabei steht X für die Antwortvariable und x für eine Ausprägung der Antwortvariablen, v indiziert eine Person, i ein Item und j eine Dimension. q_{ij} steht für die Einträge in der Designmatrix Q, θ (theta) für einen Personenparameter und σ (sigma) für einen Itemparameter, m+1 bezeichnet die Anzahl der Antwortkategorien in X, s ist ein weiterer Index zu den Antwortkategorien x.

Das Modell geht von mehreren latenten Personeneigenschaften aus. Es formuliert die Annahme, daß durch eine gewichtete Summe der loglinearen Personen-Dimensions-Parameter θ_{vj} einerseits und einen Itemparamter andererseits die Wahrscheinlichkeit einer Antwort festgelegt ist. Die Gewichte q_{ij} stammen aus einer Designmatrix Q, in der festgelegt wird, mit welchem Gewicht eine Eigenschaftsdimension j an der Antwort auf ein Item i beteiligt ist, und zwar für alle Personen v in gleicher Weise.

2.2.1 Designmatrix Q

Das folgende Beispiel zeigt eine Designmatrix Q:

		Dimension 1	Dimension2
Item	1	1	0
	2	1	0
	3	1	0
	4	0	1
	5	0	1
	6	0	1
	7	.5	.5
	8	.5	.5
	9	.5	.5

Die Gewichte stehen in jeder Zeile jeweils für ein Item und in jeder Spalte jeweils für eine Dimension. Die Designmatrix ist daher vergleichbar mit einer Ladungsmatrix der Faktorenanalysen, welche angibt, wie hoch die Items auf den Faktoren (vgl. hier die Dimensionen) "laden". Die Gewichte q_{ij} dürfen beliebige rationale Zahlen sein, sie lassen sich ebenso wie Faktorladungen als Trennschärfen der Items auf den Dimensionen interpretieren. In diesem Beispiel messen die ersten drei Items ausschließlich Dimension 1, die Items 4 bis 6 Dimension 2. Die letzten drei Items sprechen beide Dimensionen mit einer geringeren Trennschärfe als die ersten sechs Items an.

2.2.2 Personen- und Itemparameter

Das Modell sieht also für jede Person und für jede Dimension einen eigenen Parameter θ_{vj} vor. Die Personenparameter unterliegen keiner Normierungsbedingung, solange sie durch ihre Definition - die Spalten der Designmatrix - unabhängig voneinander sind. Ist dies nicht der Fall, kann die Unabhängigkeit durch eine Normierungsbedingung hergestellt werden.

Die Personenparameter sind in der Modellgleichung (Gleichung 1) additiv miteinander verknüpft und können sich daher gegenseitig kompensieren: sind zwei Fähigkeiten an der Lösung einer Aufgabe beteiligt, haben Probanden möglicherweise eine ähnliche Lösungswahrscheinlichkeit, bei denen entweder eine der beiden Fähigkeiten hoch ausgeprägt ist und die andere nicht oder bei denen beide Fähigkeiten mäßig hoch ausgeprägt sind.

Unabhängig von der Anzahl der Dimensionen beschreibt ein Parameter (je Antwortkategorie weniger eins) jedes Item: die Itemparameter beziehen sich auf die Gesamtheit der Personeneigenschaftsdimensionen. Die Beziehungen zwischen den Itemparametern und den einzelnen Personeneigenschaftsdimensionen sind ebenfalls mit der Designmatrix Q festgelegt.

Der Item(-kategorien-)parameter σ_{ix} ist eine Summe aus den alternativ zu verwendenden sog. Schwellenparametern, die die Wahrscheinlichkeit angeben, in Kategorie x zu antworten statt in der vorhergehenden Kategorie x-1. Die Schwellenparameter τ_{ix} der zu überschreitenden Schwellen (je zwischen zwei benachbarten Antwortkategorien) bis zur gegebenen Antwort x werden aufsummiert: $\sigma_{ix} = \sum_{s=0}^{x} \tau_{is}$, mit den Normierungsbedingungen $\sigma_{i0} = 0$ und $\sum_{i=1}^{I} \sum_{s=0}^{m} \sigma_{ix} = 0$. Das Konzept der Schwellenparameter stammt aus dem "partial credit" Modell von Masters (1982).

Neben dem "partial credit" oder "ordinalen" (Bezeichnung im Lehrbuch "Testtheorie Testkonstruktion" von J. Rost, 1996) Schwellenparametermodell können die Itemparameter auch gemäß dem "rating scale" Modell von Andrich (1978) und nach dem Äquidistanzmodell von Andrich (1982) geschätzt werden.

Gleichung 1 beschreibt das ordinale Modell; im "rating scale" Modell lautet der Exponentialausdruck $\exp\left(x\sum_{j=1}^{h} q_{ij}\theta_{vj} - \left(x\sigma_i + \sum_{s=1}^{x} \tau_s \right) \right)$, wobei σ_i einen Itemparameter bezeichnet und τ_s einen Schwellenparameter darstellt, der für Schwelle x bei allen Items i gleich ist. Ferner ist für eine Antwort in Kategorie "0" die Summe $\sum_{s=1}^{0} \tau_s \overset{!}{=} 0$ als Null zu verstehen.

Für das Äquidistanzmodell sieht der Exponentialausdruck wie folgt aus: $\exp\left(x\sum_{j=1}^{h} q_{ij}\theta_{vj} - \left(x\sigma_i + x(x - m_i)\delta_i \right) \right)$, σ_i stellt einen Itemlokationsparameter dar und δ_i ist ein Distanzparameter, der für alle Schwellen x eines Items i gleich ist.

Das ordinale Modell und das Äqudistanzmodell sind auch dann anwendbar, wenn für die Items eines analysierten Tests unterschiedliche Anzahlen von Antwortalternativen zur Verfügung standen.

2.2.3 Eindeutigkeit der Parameter

Die Modellparameter müssen eindeutig bestimmbar sein, um sie schätzen zu können. Dies wird durch Normierungsbedingungen erreicht.

Die Parameter für die Antwortkategorien eines Items sind voneinander abhängig, da sich die Antwortwahscheinlichkeiten zu Eins ergänzen. Bei dichotomen Items und bei mehrkategoriellen Items und ordinalem Schwellenmodell werden in MULTIRA die Parameter für die erste Antwortkategorie eines Items auf Null fixiert.

Im „Rating-Scale“ Modell werden m-1 unabhängige Schwellendistanzen geschätzt, eine Normierung ist nicht nötig. Im Äquidistanzmodell wird lediglich ein Dispersionsparameter für die Kategorien je Item geschätzt, eine Normierung ist auch hier nicht nötig. Beide Schwellenmodelle erfordern mindestens drei Antwortkategorien.

Die Personenparameter müssen unabhängig voneinander formuliert sein, d. h. die Spalten der Designmatrix dürfen nicht voneinander abhängig sein und unterliegen daher auch keiner Normierungsbedingung.

Jedoch hängen die Personenparameter in jeder Dimension von den Itemparametern ab: Addiert man zu den Parametern einer Dimension j eine beliebige Konstante c_j und addiert die Summe dieser Konstanten c_j ebenfalls zu den Itemparametern, erhält man genau die gleichen Exponenten in Gleichung 1, denn $\sum_j q_{ij}\theta_{vj} - \sigma_{ix} = \sum_j q_{ij}\left(\theta_{vj} + c_j\right) - \left(\sigma_{ix} + \sum_j q_{ij}c_j\right)$.

Um Eindeutigkeit herzustellen, muß für jede Dimension d eine Normierungsbedingung eingeführt werden. Dafür gibt es im MULTIRA-Programm drei Möglichkeiten:

Erstens kann ein Item aus jeder Dimension fixiert werden, MULTIRA fixiert die Summe der Schwellenparameter (und damit den mittleren Schwellenparameter) dieser Items auf Null. Zweitens können die Items für jede Dimension einer Summennormierung unterzogen werden. Drittens können die Personenparameter in jeder Dimensionen so normiert werden, daß ihre Summe Null ergibt (bzw. ihr Mittelwert auf Null fixiert wird).

In einigen Designmatrizen kann es abhängige Spalten geben, zum Beispiel in einem vollständigen Facetten-Design (siehe das Beispiel „Der Berliner Intelligenzstruktur Test“, S.45), wo sich die Spalten der Designmatrix Q aus jeder Facette zu Eins ergänzen und jede Facette damit ein eigenes Maß für die Gesamtfähigkeit einer Person liefert. Hier kann durch eine Facetten-Normierungsbedingung (s. Befehl Q-Matrix, S.28) der Personenparameter die Eindeutigkeit hergestellt werden. Diese Normierungsbedingung besagt, daß die Summe der Parameter einer Person in einer Facette gleich der Summe der Parameter dieser Person in den anderen Facetten gesetzt wird.

2.3 Anwendungsmöglichkeiten

Das Modell gestattet es, jede mehrdimensionale Strukturhypothese zu überprüfen, die sich in einer Designmatrix Q beschreiben läßt.

An Gleichung 1 wird eine Ähnlichkeit zwischen dem MULTIRA-Modell und der Faktorenanalyse deutlich. Die Antwortwahrscheinlichkeit einer Person für ein Item hängt neben anderem von einer Linearkombination über die gewichteten Personeneigenschaftsausprägungen ab. Formal besteht damit eine Ähnlichkeit zur Faktorenanalyse, wo der Personenmeßwert durch eine gewichtete Summe der Faktorwerte ausgedrückt wird, wenngleich in der Faktorenanalyse keine Itemantworten modelliert werden. Da für eine Analyse mit dem MULTIRA-Modell zunächst eine Hypothese über die mehrdimensionale Komponentenstruktur der Items in Form einer Designmatrix Q formuliert werden muß, ist das Modell ist mit einer konfirmatorischen Faktorenanalyse vergleichbar.

Die Struktur der Interkorrelationen der geschätzten Personenparameter ist nicht durch die Modellstruktur festgelegt, sondern hängt lediglich von der Designmatrix Q und den Daten ab. Die Interkorrelationsstruktur der Personendimensionsparameter stellt daher ein interpretierbares Ergebnis dar.

Im folgenden werden zwei Modelle dargestellt, die sich als Spezialfall des MULTIRA-Modelles darstellen lassen.

Zunächst ein Modell von Klauer und Sydow (), welches Lernen während der Testbearbeitung modelliert. Das Modell in Gleichung 2 sieht neben einem Personenfähigkeitsparameter θ_v einen weiteren Personenfähigkeitsparameter δ_v vor, welcher mit jedem weiteren Item ein stärkeres Gewicht erhält. Außerdem stellt ein Itemparameter die Eigenheiten jedes Items dar.

Gleichung 2 $$p(X_{vi} = x) = \frac{\exp(x\theta_v + x(i-1)\delta_v + \sigma_{ix})}{\sum_{s=0}^{m} \exp(s\theta_v + s(i-1)\delta_v + \sigma_{is})}$$ Lerngewinn-Modell

Mit der folgenden Designmatrix läßt sich dieses Modell als Spezialfall des MULTIRA-Modells darstellen.

	Grundfähigkeit	Lernfähigkeit
1	1	0
2	1	0.1

	3	1	0.2
I =	3	1	0.3
	4	1	0.4
	5	1	0.5
	...	...	...
	20	1	1.9

Abbildung 1 Q-Matrix zum Lerngewinn-Modell

Spalte 1 enthält gleich hohe Gewichte, damit repräsentiert die Komponente 1 den Personenparameter θ_v. In Spalte 2 nehmen die Gewichte von Item zu Item um einen bestimmten Betrag zu. Damit stellt Komponente 2 den Lernfähigkeitsparameter δ_v dar. Mit der Parameterschätzung erhält man für jede Person getrennt die Grundfähigkeit und die Lernfähigkeit.

Einen weiteren Spezialfall des Modells stellen Fragebögen dar, die von unterschiedlichen Facetten in Personenvariablen ausgehen. Die Beantwortung eines Items hängt von dabei von einer (aus mehrenen) Personeneigenschaft je Facette ab. Ein Beispiel wäre ein Intelligenztest, der von verschiedenen Operationen ausgeht, die nur jeweils im Zusammenhang mit bestimmten Inhalten gemessen werden können. Die folgende Desgnmatrix stellt die vier möglichen Itemtypen dar, die sich ergeben, wenn man zwei Operationen und zwei Inhalte annimmt.

Item(-typ)	Operation 1	Operation2	Inhalt 1	Inhalt 2
1	1	0	1	0
2	1	0	0	1
3	0	1	1	0
4	0	1	0	1

Abbildung 2 Q-Matrix zum Facetten-Modell

Um dieses Facettendesign zu illustrieren, wird in Kapitel 6 das Berliner Intelligenzstruktur-Modell vorgestellt, welches in der Facetten "Operationen" vier Dimensionen unterscheidet und die Facette "Inhalte" in drei Dimensionen unterteilt.

2.4 Parameterschätzungen

Die Itemparameter können auf zwei Arten geschätzt werden, durch die Maximierung der bedingten Likelihood (**CML**) oder durch die Maximierung der unbedingten Likelihood (**UML**). Die bedingten Schätzer sind unverzerrt gegenüber den unbedingten Schätzern und sind von daher vorzuziehen. Jedoch unterliegt die bedingte Schätzung gewissen Einschränkungen, so daß die Verwendung der unbedingten Schätzer sinnvoll sein kann.

Die Scores, nach denen die Itemparameterschätzung bedingt wird, bestehen im mehrdimensionalen Modell aus einem d-dimensionalen Feld. Die Zahl der unterscheidbaren Scores steigt mit zunehmeder Zahl von Dimensionen rapide an: Bei zehn dichotomen Items in einer Dimension gibt es elf unterschiedliche Scores, bei zehn dichotomen Items in je zwei Dimensionen gibt es 121 unterschiedliche Scores, bei drei Dimensionen mit je zehn dichotomen Items sind bereits 1331 Scores zu unterscheiden. Diese Zahlenbeispiel gilt nur dann, wenn die q-Gewichte in der Designmatrix nur aus Einsen und Nullen bestehen. Läßt man unterschiedliche (ganzzahlige) Gewichte zu, erhöht sich die Zahl der Scores noch stärker.

Das hat bei einer festen Stichprobengröße zur Folge, daß die Häufigkeiten der beobachteten Scores kleiner werden und viele Scores gar nicht mehr beobachtet werden. Außerdem steigt der Rechenaufwand ebenso rapide, so daß die maximale Zahl von Dimensionen auch eine Zeitfrage wird.

Im unbedingten Maximum-Likelihood Verfahren unterliegen die Itemparameterschätzungen einem Bias, für den bisher nur heuristische Korrekturterme vorgeschlagen wurden. Dieser Bias fällt jedoch nur ins Gewicht, wenn die Verteilung der Itemparameter asymmetrisch ist (Wollenberg et al., 1988) und wird für größere Zahlen von Items kleiner (Rost, 1983). Von daher kann dieser Bias in Kauf genommen werden, wenn UML-Schätzer verwendet werden.

Die Schätzung der Personenparameter des MULTIRA-Modells beruht die Maximierung der unbedingten Likelihood. Zur Verbesserung der Schätzung wird die weighted Likelihood nach Warm (1989), eine modifizierte unbedingte Likelihood, maximiert.

Im vereinfachten Newton-Raphson-Algorithmus (z. B. in Rost, 1988) wird die Likelihood für jeden Modellparameter einzeln maximiert, anstatt über den Vektor aller Parameter gleichzeitig. Dieses Verfahren hat sich häufig im Zusammenhang mit Rasch-Modellen bewährt und wird daher auch im vorliegenden Modell verwendet. Wenn die Likelihood maximal wird, liegt in deren erster Ableitung eine Nullstelle vor. Für jeden Modellparameter wird iterativ die Nullstelle der ersten partiellen Ableitung der Likelihood nach diesem Parameter bestimmt.

Warm (1989) untersucht den Bias dieser uL-Schätzer und modifiziert die unbedingte Likelihood uL um einen Term, der nach dem Erwartungswert des Bias der Personenparameter-schätzungen abgeleitet ist. Maximiert man diese weighted Likelihood wL, erhält man genauere Schätzer (wL-Schätzer) als die uL-Schätzer für die Personenparameter.

2.5 Modellgeltungskontrolle

In diesem Abschnitt werden kurz die mit MULTIRA möglichen Verfahren für Modellgeltungskontrollen aufgezählt. Eine ausführlichere Darstellung gibt Carstensen (2000, Kapitel 4); in den Beispielen in Kapitel 6.3 dieses Handbuchs wird die Anwendung der Modellgeltungstests außerdem demonstriert.

Verwendet man CML-Schätzungen, so stehen die bedingte Likelihood bzw. eine marginale Form der bedingten Likelihood zu Verfügung und es können Modellvergleiche durch **Likelihood-Ratio** Tests durchgeführt werden. Außerdem berechnet MULTIRA die **Informationskriterien** AIC, BIC und CAIC.

In Verbindung mit UML-Schätzern und der unbedingten Likelihood müssen die Eigenschaften dieser Modellfitmaße erst noch untersucht werden und können daher nicht verwendet werden.

Ein Verfahren zur Modellgeltungskontrolle, welches nicht auf der Likelihood basiert, besteht in einem sogenannten **Bootstrap**-Verfahren bezüglich einer Abweichungsstatistik zwischen den beobachteten Patternhäufigkeiten und den unter Modellgeltung zu erwartenden Patternhäufigkeiten. Um eine Prüfverteilung für die eben beschriebene χ^2 -Statistik zu erhalten ist es möglich, eine Reihe von Datensätzen anhand der geschätzten Modellparameter zu resimulieren, so daß auf diese Weise empirisch die Verteilung der Abweichungsstatistik unter der Bedingung dieser Parameter angenähert wird. Näheres zum Bootstrap-Verfahren bei von Davier (1997).

Kann von Modellgeltung ausgegangen werden, so läßt sich an den Residuen der Itemantworten ablesen, wie gut die einzelnen Items zum Modell passen. Häufig verwendete Statistiken dafür sind die sogenannten Maße **Infit** und **Outfit** (die als unweighted und weighted Mean Square von Wright und Masters (1982) beschrieben werden). Beide Fitmaße aggregieren für jedes Item, wieviele (unter dem geschätzten Modell) „unerwartete“ Antworten beobachtet wurden. Beide Fitmaße haben einen Erwartungswert von Eins; weichen die Fit-Werte signifikant von Eins ab, paßt das betreffende Item nicht zu den Daten. Ein zu hoher Wert indiziert eine niedrige Trennschärfe des Items (zu viele unerwartete Antworten), ein zu niedriger Wert bedeutet eine zu hohe Trennschärfe (zu wenig unerwartete Antworten).

Reckase (1998) schlägt vor, die **Kovarianz-Matrizen** der Itemresiduen heranzuziehen, um die Qualität der Modellanpassung zu abzulesen. Hierfür gibt es keine statistischen Tests, sondern es sind einzelne Kovarianzen zu vergleichen.

Einen Test zur Modellgeltung des Spezialfalles des MultiRa-Modells in Gleichung 5, wo keine separaten Itemparameter geschätzt werden (siehe unten) beschreibt Stegelmann (1983). Bei diesem Test handelt es sich um einen Likelihoodquotiententest, der die Personenhomogenität anhand von mehreren Personengruppen untersucht:

Teilt man die Stichprobe in G disjunkte Personengruppen, indiziert mit g, und schätzt die Parameter für diese Gruppen getrennt und und für die Gesamtstichprobe, indiziert mit T, so ist $-2\ln\frac{uL_T}{\prod_g uL_g}$ asymptotisch χ^2-verteilt mit (G-1)*N*J Freiheitsgraden, wobei N die Anzahl der Personen in T und J die Anzahl der Dimensionen darstellt.

2.6 Relation zu anderen Testmodellen

2.6.1 Herleitung

Zunächst wird kurz vorgeführt, wie sich das MULTIRA-Modell aus dem einfachen logistischen Modell von Rasch (1960) (Gleichung 3) und mit dem Ansatz von Item-Komponenten-Modellen herleiten läßt.

Gleichung 3 $p(x=1|v,i)=\frac{\exp(\theta_v-\sigma_i)}{1+\exp(\theta_v-\sigma_i)}$, einfaches Rasch-Modell

Zunächst tritt hier eine gewichtete Summe aus dimensionsspezifischen Personenparametern und Itemparametern an die Stelle der Summe aus Personenparameter und Itemparameter im einfachen Rasch-Modell, jeweils gewichtet durch den entsprechenden Koeffizienten aus der Designmatrix Q. Das resultierende Modell (Gleichung 4) enthält folglich N x J Personen(-komponenten-)parameter und I x J Item(-komponenten-)parameter bei N Versuchspersonen, I Items und J Komponenten.

Gleichung 4 $p(X=1|v,i)=\frac{\exp\left(\sum_{j=1}^{J}\left(q_{ij}\left(\theta_{vj}-\eta_{ij}\right)\right)\right)}{1+\exp\left(\sum_{j=1}^{J}\left(q_{ij}\left(\theta_{vj}-\eta_{ij}\right)\right)\right)}$, allgemeines Itemkomponentenmodell

Es stellt eine sehr allgemeine Struktur dar, die ohne eine Vielzahl von Nebenbedingungen nicht identifizierbar ist und so viele Parameter enthält, daß es schwer sein wird, alle Parameter des Modells zu interpretieren.

Nimmt man an, daß die Personenkomponentenparameter komponentenunabhängig sind, verbleiben lediglich n Personenparameter in der Modellstruktur; man erhält ein eindimensionales Modell. Restringiert man die Itemkomponentenparameter durch die Annahme, daß diese Parameter itemunabhängig sind, kann man den Index i fortlassen. Die verbleibenden Parameter η_j sind komponentenspezifische Basisparameter, die für alle Items gelten, und aus denen sich die Itemschwierigkeiten zusammensetzen. Diese beiden Annahmen führen zum linearlogistischen Testmodell (LLTM von Fischer, 1974), Gleichung 9.

Um jedoch ein mehrdimensionales Modell zu erhalten, restringierte Stegelmann (1983) lediglich die Itemkomponentenparameter in der eben beschriebenen Weise. Nach der Reparametrisierung $\theta_{vj}^{*} = \theta_{vj} + \eta_j$ resultiert die Modellstruktur in Gleichung 5, die lediglich Personenkomponentenparameter enthält und keine Itemeigenschaften modelliert.

Gleichung 5 $p(x = X | v, i) = \dfrac{\exp\left(x \sum_{j=1}^{J} q_{ij} \theta_{vj}^{*} \right)}{\sum_{s=0}^{m} \exp\left(s \sum_{j=1}^{J} q_{ij} \theta_{vj}^{*} \right)}$, Stegelmann Modell

Restringiert man statt dessen die Itemkomponentenparamter durch die Annahme, sie komponentenunabhängig, d. h., sie stellen nur Itemeigenschaften dar, so folgt durch die Reparametrisierung $\overset{*}{\sigma}_i = \left(\sum_{j=1}^{J} q_{ij} \right) \sigma_i$ das Modell (Gleichung 6), welches neben Personenkomponentenparametern komponentenunabhängige Itemparameter enthält.

Gleichung 6 $p(x = 1 | v, i) = \dfrac{\exp\left(\sum_{j=1}^{J} q_{ij} \theta_{vj} - \sigma_i \right)}{1 + \exp\left(\sum_{j=1}^{J} q_{ij} \theta_{vj} - \sigma_i \right)}$, Mehrdimensionales-Komponenten- Rasch-Modell für dichotome Daten

Dieses Modell ist bereits das MULTIRA-Modell im Falle dichotomer Daten. Läßt man mehr als zwei Antwortkategorien zu, ergibt sich mit der bereits beschriebenen Parameterisierung durch Schwellenparameter das MULTIRA-Modell für ordinale Daten.

2.6.2 MRCML & ConQuest

Das MRCML von Wilson et al. ist ein allgemeineres Modell als MULTIRA, es sieht zum Beispiel noch eine Zerlegung der Itemparameter in Komponenten vor, die eine andere Struktur aufweisen muß, als die Struktur der Personenkomponenten. Das Programm ConQuest zur Schätzung der Parameter des MRCML beruht auf einem marginal Maximum Likelihood Verfahren und bietet auch Schätzungen der Verteilungsparameter der Personenfähigkeitsverteilung an.

Das MULTIRA-Programm ist in der Lage größere Zahlen von Dimensionen zu verarbeiten als ConQuest. Ferner sei darauf hingewiesen, daß das MultiRa-Modell weniger komplex ist und die Bedienung des Menügesteuerten Windowsprogramms MULTIRA möglicherweise leichter fällt.

2.6.3 Das One-Parameter Logistic Model OPLM

Das One-parameter-logistic-Model OPLM von Verhelst, Glas und Verstralen (1994) ist identisch mit dem MULTIRA-Modell, wenn man ein eindimensionales Modell mit unterschiedlichen Itemtrennschärfen spezifiert. Die Designmatrix Q enthält dann eine Spalte, deren Einträge unterschiedliche Rationalzahlen annehmen kann, die Gewichte stellen die Itemtrennschärfen dar. Die Parameter des OPLM werden mit dem bedingten Maximum Likelihood Verfahren geschätzt.

In der Item-Response-Theorie haben die Trennschärfen eine besondere Bedeutung, da mit unterschiedlichen Itemtrennschärfen sich überschneidende Item-Charakteristik-Kurven angenommen werden. Mit sich überschneidenden ICC folgt, daß die Rangreihenfolge der Items sich für Personen unterschiedlichen Fähigkeitsgrades ändern kann. Dies steht im Widerspruch zur Eindimensionalitätsannahme im Rasch-Modell. Im "6.1 Eine, zwei oder vier Dimensionen ? Ein Physikleistungstest" in Kapitel 6 (S. 45) werden unterschiedliche Wege dargestellt, Mehrdimensionalität zu modellieren.

3. Weitere Testmodelle

Das Programm MULTIRA bietet neben dem MultiRa-Algorithmus weitere Algorithmen an. Dazu gehört der Algorithmus MKAT zum mehrkategoriellen Modell von G. Rasch, ein Algorithmus zum Martin-Löf-Test und der Algorithmus zum Linear logistischen Testmodell LLTM für dichotome Daten (G. Fischer, 1974).

3.1 MKAT

Der Algorithmus MKAT schätzt die Parameter des mehrkategoriellen logistischen Modells von G. Rasch (1961). Das Modell modelliert die Annahme, daß "die Auswahl" einer Anwortkategorie für ein Item von je einer latenten Variable für jede Antwortkategorie abhängt. Die Antwort einer Person gibt also Aufschluß über die Ausprägung dieser kategorienspezifischen latenten Variablen relativ zueinander. Die Modellparameter sind folglich ipsative Meßwerte. Die Antwortwahrscheinlichkeiten nach diesem Modell beschreibt die folgende Gleichung 7.

Gleichung 7 $$p(X_{vi} = x) = \frac{\exp(\theta_{vx} - \sigma_{ix})}{\sum_{s=0}^{m} \exp(\theta_{vs} - \sigma_{is})}$$, mehrkategorielles Rasch-Modell

Der Algorithmus MKAT von J. Rost verwendet eine bedingte Maximum Likelihood Methode zur Parameterschätzung und wurde unter Verwendung von Programmteilen von Scheiblechner, abgedruckt in Fischer, 1974, S. 555ff., geschrieben. Eine Besonderheit liegt in der Normierungsbedingung für die Personenparameter: Durch eine Summennormierung

$$\sum_{x=0}^{m} \theta_{vx} = 0$$

erhält man für jede Antwortkategorie einen Parameter.

Weitere Informationen finden sich im Kapitel 3.2.2 im Lehrbuch "Testtheorie - Testkonstruktion" von J. Rost (1996).

3.2 MLT

Der Martin-Löf-Test ist ein Signifikanztest für Itemhomogenität gemäß dem Rasch-Modell für dichotome Daten. Die Items eines Tests werden nach einer existierenden Hypothese in zwei möglichst heterogene Itemgruppen aufgeteilt und für beide Testteile getrennt die Parameter unter Maximierung der bedingten Likelihood geschätzt. Der Test basiert auf den Häufigkeiten n_{rs}, mit denen die Personen im einen Testteil einen Score r und im anderen einen

Score s erhalten haben. Die Prüfgröße ist ein modifizierter Likelihoodquotiententest, dargestellt in Gleichung 8, wobei k die Zahl der Items im Gesamttest ist, k_1 die Zahl der Items in Testteil 1 und k_2 im Testteil 2, cL_0 die bedingte Likelihood des Gesamttests, cL_1 und cL_2 die der beiden Testeile. Der Score einer Person im Gesamttest wird mit n_r bezeichnet, der Score r im ersten Testteil und s im zweiten Testteil mit n_{rs}.

Gleichung 8 $$-2\log\frac{\prod_{r=0}^{k}\left(\frac{n_r}{N}\right)^{n_r} cL_0}{\prod_{r=0}^{k_1}\prod_{s=0}^{k_2}\left(\frac{n_{rs}}{N}\right)^{n_{rs}} cL_1 cL_2}$$, Prüfgröße im Martin-Löf-Test MLT

Wird die Prüfgröße signifikant größer als unter der χ^2-Verteilung mit $df = k_1 k_2 - 1$ zu erwarten, muß die Annahme homogener Testteile fallengelassen werden.

Weitere Informationen finden sich im Kapitel 5.3.2 im Lehrbuch "Testtheorie - Testkonstruktion" von J. Rost (1996).

3.3 LLTM

Das Linear Logistische Testmodell von Fischer (1972) ist ein Itemkomponenten-Modell, welches von einer latenten Variable ausgeht und die Itemschwierigkeiten mithilfe einer Designmatrix auf sogenannte Basisparameter zurückführt (s.o.). Über die Modellgeltung des Rasch-Modells hinaus müssen sich die Itemschwierigkeiten aus Kombinationen der Basisparameter darstellen lassen. Es werden somit Modelle mit weniger Modellparametern als im unrestringierten Rasch-Modell formuliert. Die Lösungswahrscheinlichkeiten ergeben sich wie in Gleichung 9 dargestellt, wobei j die Basisparameter η_j indiziert.

Gleichung 9 $$p(x_{vi}=1)=\frac{\exp\left(\theta_v-\sum_j q_{ij}\eta_j + c\right)}{1+\exp\left(\theta_v-\sum_j q_{ij}\eta_j + c\right)}$$, Linear logistisches Testmodell LLTM

Eingehendere Beschreibungen des Modells finden sich bei Fischer, zum Beispiel im Lehrbuch (1974).

4. Die Bedienung der Benutzeroberfläche

Zunächst eine Übersicht über MULTIRA's Menüs und Befehle:

In dem Menü "**Datei**" (S. 24) finden sich die Befehle **Datensatz**, **Q-Matrix**, ein Text Editor mit den Befehlen **Neu**, **Öffnen**, **Speichern** & **Speichern unter** und **Beenden**.

Im Menü "**Bearbeiten**" (S. 29) finden sich die Befehle **Kopieren**, **Ausschneiden**, **Einfügen**, **Voreinstellungen** (S. 29) und **Stapellauf** (S. 29).

Unter dem Menü „**Programm**" (S. 30) finden sich die Befehle **MULTIRA** (S. 31)**, MKAT** (S. 34)**, MLT** (S. 35) und **LLTM** (S. 35). Jeder dieser Befehle steht für eine Rechenroutine; die Erläuterung der Programme erfolgt in der genannten Reihenfolge.

Das Menü "**Fenster**" (S. 37) enthält die Befehle **Daten**, **Ergebnisse** und **Anordnen**.

Das Menü „**?** " (S. 37) beinhaltet die Befehle **Hilfe** und **Über MULTIRA**.

4.1 Das Menü „Datei"

In dem Menü **„Datei"** finden sich die Befehle **Datensatz** (S. 24), **Q-Matrix** (S. 27), **Neu**, **Öffnen**, **Speichern**, **Speichern unter** (S.28) und **Beenden** (S. 29).

4.1.1 Datensatz

Vorweg: Die Auswahl eines Datensatzes in MULTIRA beinhaltet die Definiton der beobachteten Variablen und ggf. Rekodierungen. Der ausgewählte Datensatz wird rekodiert gespeichert und steht anschließend für Analysen mit einem der von MULTIRA angebotenen Programme zur Verfügung.

Hinweis: Alle Definitionen zu einem Datensatz werden in einer Datei mit gleichem Namen und der Endung >.ini< abgespeichert; der rekodierte Datensatz wird mit gleichem Namen und der Endung >.rec< abgespeichert.

Die Auswahl des Befehls **Datensatz** Befehls führt zum Dialog-Fenster *"Datensatz auswählen"*. Hier können der Dateiname des zu öffnenden Datensatzes und der entsprechende Pfad angegeben werden. Die zunächst vorgegebene Dateitypenendung *.dat kann verändert werden. Das Öffnen eines ausgewählten Datensatzes führt zum Fenster *"Kennwerte der Daten"*. Auf die hier verfügbaren Optionen geht der folgende Abschnitt ein. Gleichzeitig wird ein Fenster *"Data- \>Pfad, Name des Datensatzes<"* geöffnet, in dem der aktuell geöffnete Datensatz sichtbar ist.

4.1.1.1 Itempositionen

Die Anzahl der Items sowie die Spaltenbreite der Items können in die dafür vorgesehenen Kontrollfelder im Fenster *"Kennwerte der Daten"* eingegeben werden.

Grundsätzlich erkennt das Programm nur Datensätze mit fester Breite („Fixed-Format“) aller Itemkodierungen. Es sind bis zu acht Zeichen je Item zulässig,. Bei nominal kodierten Item-Antworten (siehe unten) dürfen beliebige Zeichen verwendet werden.

Die Items werden durch Nummern von 1 aufsteigend bezeichnet und nicht durch andere Itembezeichnungen (Labels). Bei Einstellung einer Spaltenbreite von "1" ist entsprechend die Anzahl der Items gleich der Anzahl der Spalten im Datensatz. Die Spalten, in denen die Items stehen, sind frei wählbar. Die Auswahl kann in dem Kontrollfeld „Spalten der Items“ vorgenommen werden. Die Angaben der Spalten müssen dabei durch ein Leerzeichen getrennt werden.

Bei Auswahl der Option „Auto-Format“ mittels des entsprechenden Kontrollkästchens werden die Spalten, in denen die Items gesucht werden, automatisch in aufsteigender Reihenfolge mit der vorgegebenen Spaltenbreite in das Textfeld „Spalten der Items“ eingetragen. Stehen die Itemkodierungen in anderen Spalten, sollte das Kontrollkästchen "Auto-Format" ausgeschaltet werden.

4.1.1.2 Antwortkategorien

Im nächsten Kasten kann mit den entsprechenden Kontrollkästchen angeben werden, ob nominal kodierte Item-Antworten, ordinal kodierte Antwortkategorien oder Intervalldaten vorliegen.

Bei nominal kodierten Antworten muß die Anzahl der Antwortkategorien (2-15) angegeben werden, es müssen in allen Items gleich viele Antwortkategorien zur Verfügung gestanden haben. In das Kontrollfeld „Kategorien-Codes“ können die Kodierungen der Kategorien, durch eine Leerstelle getrennt, angegeben werden; dabei dürfen Ziffern und Buchstaben verwendet werden.

Mit der Option „Auto-Format“ werden die Kategorien-Codes automatisch für ordinale Kategorienkodierungen vorgeschlagen. Sollen andere Kodierungen verwendet werden, muß das Kontrollkästchen "Auto-Format" ausgeschaltet werden.

Bei ordinal kodierten Antwortkategorien ist die niedrigste und die höchste Kategorie anzugeben. Werte, die ausserhalb des Intervalls liegen, werden als "missing values" behandelt. Die

Daten werden so rekodiert, daß die niedrigste angegebene Kategorie auf "Null" gesetzt wird, daher ist die niedrigste Kategorie genau anzugeben.

Sollen Intervalldaten eingelesen werden, muß die Option "ordinal kodierte Antwortkategorien gewählt werden und die Daten müssen rekategorisiert werden (siehe unten).

Der MultiRa-Algorithmus rechnet mit bis zu 15 Antwortkategorien je Item, wobei unterschiedliche Kategorienzahlen je Item verwendet werden dürfen. Das LLTM erfordert dichotome Daten und MKAT erfordert nominale Daten mit gleicher Anzahl von Antwortkategorien je Item.

4.1.1.3 Rekategorisieren

Mit dem Kontrollfeld „Verteilt auf (2-150) Kategorien“ lassen sich die Antworten auf eine gewünschte Anzahl von Kategorien verteilen. Dabei werden die neuen Kategorien mit gleichen Abständen über (genau) das gesamte Intervall der alten (und nicht als Bezeichner von fehlenden Werten ausgeschlossenen) Antwortkategorien verteilt. Es kann ausgewählt werden, ob für alle Items auf die gleiche Zahl von Kategorien erzeugt werden soll (Schaltfeld „genau“) oder ob für jedes Item separat eine Zahl von Kategorien festgelegt werden soll (Schaltfeld „maximal“). Im letzteren Falle wird nach dem einlesen der Daten ein Dialogfenster geöffnet, welches für jedes Item die Zahl der beobachteten kategorien und die neu festzulegende zahl anzeigt.

Hinweis: Diese Option ermöglicht auch das Kategorisieren von Intervalldaten.

4.1.1.4 Fehlende Werte

Hier kann ausgewählt werden, wie fehlende Werte behandelt werden sollen. Personen mit fehlenden Werten können komplett ausgeschlossen werden ("Fall ausschließen") oder fehlende Werte können durch korrigierte Itemmittelwerte ersetzt werden ("Ersetzen durch korrigierte Itemmittelwerte"). Dabei wird die mittlere Itemantwort für das betreffende Item $\overline{x}_i$ mit dem Verhältnis aus Personenscore der betreffenden Person zum mittlerem Personenscore $\frac{r_v}{\overline{r}}$ multipliziert: $x_{v,i,neu} = \left(\overline{x}_i \frac{r_v}{\overline{r}} \right)$ und anschließend auf eine ganze Zahl gerundet. Das Ersetzen folgt grob einem eindimensionalen Antwort-Modell und führt neben einer Veränderung des Datensatzes auch zu einer Beeinflussung der Modellgeltung. Es hat sich in Simulationsstudien als robuste Methode erwiesen (Huismann & Molenaar,1997).

Außerdem können fehlende Werte auf eine (für alle Items) gültige Kategorie nach Wahl umkodiert werden ("umkodieren auf ...").

4.1.1.5 Datensatz prüfen

Sind die Angaben vollständig, kann der Datensatz geprüft werden. Nach dem Einlesen der Daten informiert das Fenster *"Eigenschaften des Datensatzes"* über

- die Anzahl gefundener Personen (bzw. Zeilen),
- die Anzahl der Personen mit fehlenden Werten und
- eine Statistik über die Zahl der Antworten je Antwortkategorie für alle Items.

Stimmen diese Angaben mit dem Datensatz überein, wird durch Betätigen der Schaltfläche "Fortfahren" der Datensatz rekodiert und gespeichert. Er steht jetzt für Analysen zur Verfügung.

Der rekodierte Datensatz wird unter >Dateiname.rec<< gespeichert, in der Datei >Dateiname.ini< werden die Angaben über diesen Datensatz gespeichert. Wurde ein Datensatz schon einmal geöffnet und rekodiert abgespeichert, können diese Einstellungen durch betätigen der Schaltfläche "Akzeptieren" übernommen werden und der Datensatz für Analysen zur Verfügung gestellt werden.

Hinweis: Alle erforderlichen Einstellungen müssen für jeden neu geöffneten Datensatz eingegeben werden. Angezeigt werden die Einstellungen für den zuletzt geöffneten Datensatz bzw. willkürliche Voreinstellungen.

4.1.2 Q-Matrix

Erstellung der Designmatrix Q

Wählt man aus dem Menü "**Datei**" den Befehl **Q-Matrix**, führt dies zum Dialogfenster *"Q-Matrix wählen"*. Hier können der Dateiname der zu öffnenden Q-Matrix-Datei und der entsprechende Pfad angegeben werden. Das Programm erkennt zunächst nur Q-Matrix-Dateien mit der Dateitypenendung *.qmx. Diese kann jedoch bei Bedarf geändert werden. Q-Matrix-Dateien müssen im Textformat geschrieben werden.

Das Öffnen einer Q-Matrix-Datei führt zum Fenster *"Q-Matrix \>Pfad, Name der Q-Matrix-Datei<"*. Bei erstmaliger Benutzung von MULTIRA kann die Datei "beispiel.qmx" geöffnet werden. Die in dieser Datei enthaltene Beispieldesignmatrix kann verändert und mit der Schaltfläche "Speichern unter“ unter einem neuen Namen gespeichert werden.

Über der Beispielmatrix finden sich Angaben zum erforderlichen Format der Designmatrix Q. Generall können außerhalb der Matrix selber Kommentare in die Datei eingefügt werden.

Nach einer Abschnittsmarkierung "[qmatrix]" folgt unmittelbar die Designmatrix Q. Nach der Designmatrix Q muß eine leere Abschnittsmarkierung "[]" stehen. Die Koeffizienten können wahlweise durch ein Leerzeichen oder Semikolon getrennt werden; Dezimalstellen können entweder durch Punkt oder durch Kommata getrennt werden.

Soll die Facetten-Normierungsbedingung (s. MultiRa-Modell, S.14) der Personenparameter im MultiRa-Modell verwendet werden, muß die Abschnittsmarkierung nach der Matrix ein Schlüsselwort enthalten, dessen erste fünf Buchstaben "facet" in Groß- oder Kleinschreibung lauten. Die Zeile danach muß für jede Dimension eine ganze Zahl enthalten, die die Zugehörigkeit zu einer Facette angibt. Im folgenden Beispiel sind die letzten beiden Zeilen für ein vierdimensionales Modell mit zwei Facetten angegeben:

```
[Facetten]

1 1 2 2
```

Hinweis: Das letzte Zeichen in der zweiten (und gleichzeitig letzten) Zeile muß von einem Leerzeichen oder einer weiteren Leerzeile gefolgt sein, es darf nicht das letzte Zeichen in der Datei sein.

4.1.3 Job laden

Hier kann ein zuvor gespeicherter Job erneut geladen werden. Der Datensatz und alle Definitionen werden eingelesen und das MultiRa-Programmmenu geöffnet.

4.1.4 Text-Editor

Die Befehle Neu, Öffnen, Speichern und Speichern unter gehören zu einem einfachen Texteditor im Hauptfenster von MULTIRA. Eine in einem Editorfenster geöffnete Datei kann verändert und gespeichert werden. Es steht das Menü "Bearbeiten" zur Verfügung, dessen Funktionen auch mit den üblichen Tastenkombinationen zu verwenden sind.

Der Befehl **Neu** führt zu einem neuen leeren Editor-Textfenster.

Neben dem Editor-Fenster gibt es in MULTIRA noch weitere Textfenstertypen, nämlich das Datensatz- und Ergebnisfenster, die nicht verändert werden können, und das Q-Matrix- ‚das Voreinstellungs- und ferner das Stapellauf-Fenster, in denen die Bearbeiungsfunktionen ebenfalls zur Verfügung stehen.

Der Befehl **Öffnen** führt zum Dialogfenster *„Öffnen“*. Es können beliebige Dateien in einem Editor-Fenster geöffnet werden.

Mit dem Befehl **Speichern** können Dateien in aktuell geöffneten Textfenstern gespeichert werden.

Mit dem Befehl **Speichern unter** lassen sich Dateien in aktuell geöffneten Textfenstern unter einem anderen als dem aktuell bestehenden Dateinamen speichern.

4.1.5 Beenden

Beendet MULTIRA.

4.2 Das Menü „Bearbeiten“

Im Menü „Bearbeiten“ finden sich die Befehle **Kopieren**, **Ausschneiden**, **Einfügen, Voreinstellungen** (S. 29) und **Stapellauf** (S. 29).

Die Befehle **Kopieren**, **Ausschneiden** und **Einfügen** sind mit den in Windows-Programmen üblichen Funktionen in Textfenstern verbunden und können auch über die Tastenkombinationen Strg-C, Strg-X und Strg-V bzw. Strg-Einfg, Umschalt-Entf und Umschalt-Einfg aufgerufen werden.

4.2.1 Voreinstellungen

Der Befehl **Voreinstellungen** führt zum Fenster *„Voreinstellungen“*. Fenster verändert werden müssen, sondern als benutzerdefinierte Voreinstellungen in bestimmten Kontrollfeldern verwendet werden.

4.2.2 Stapellauf

Ein Stapellauf (engl.: „batch run“) ist das aufeinanderfolgende Abarbeiten mehrerer, vorher gespeicherter Jobs. Ein **Job** besteht aus allen notwendigen Angaben über den Datensatz und das Modell für eine Parameterschätzung oder einen Test.

Hinweis: „Job“-Dateien werden in Textformat mit der Endung >.job< gespeichert. Zur Zeit können nur Multira-Jobs gespeichert werden (S. 34).

Der Befehl **Stapellauf** führt zum Fenster *„Einen Stapellauf (Batch-Run) editieren“*. Das Fenster beinhaltet eine Menüleiste mit den Menüs „**Datei**“, „**Bearbeiten**“ und „**Ende**“ sowie die Befehle „**Job hinzufügen**“ und „**Start**“.

Das Menü **„Datei“** im Fenster *„Einen Stapellauf (Batch-Run) editieren“* enthält die Befehle **Stapel laden** und **Stapel speichern unter**. Hier können Stapeldateien, nachdem sie erstellt wurden, abgespeichert und später wieder aufgerufen werden. Der Befehl **Stapel laden** führt zum Fenster *„Stapeldatei auswählen“.* Anzugeben sind hier der Dateiname einer zu öffnenden Stapeldatei und der entsprechende Pfad. Das Programm erkennt zunächst nur Stapeldateien mit der Dateitypenendung *.bat. Die Dateitypenendung kann jedoch verändert werden. Der Befehl **Stapel speichern unter** führt zu einem Dialog-Fenster *„Stapel speichern unter“*. Hier lassen sich aktuell geöffnete Stapeldateien unter einem anderen als dem aktuell bestehenden Dateinamen speichern.

Das Menü **„Bearbeiten“** im Fenster *„Einen Stapellauf (batch-run) editieren“* enthält die Befehle **Kopieren, Ausschneiden** und **Einfügen.** Diese Befehle sind mit den in Windows-Programmen üblichen Editier-Funktionen verbunden. Der Befehl **Job hinzufügen** im Fenster *„Einen Stapellauf (batch-run) editieren“* führt zum Fenster *„Job auswählen“*. Hier können vorher definierte Jobs in den aktuellen Stapel aufgenommen werden.

Mit dem Befehl **Start** aus der Menüleiste kann der im Fenster angezeigte Stapel aus zuvor ausgewählten Jobs ausgeführt werden.

Das Menü **„Ende“** im Fenster *„Einen Stapellauf (batch-run) editieren“* enthält schließlich mit den Befehlen **mit speichern** und **ohne speichern** die Option, den aktuellen Stapel vor dem Schließen des Fensters zu speichern.

4.3 Das Menü „Programm“

Unter dem Menü **„Programm“** finden sich die Befehle **MULTIRA** (S. 31), **MKAT** (S. 34), **MLT** (S. 35) und **LLTM** (S. 35). Jeder dieser Befehle steht für eine Rechenroutine; die Erläuterung der Programme erfolgt in der genannten Reihenfolge.

4.3.1 MULTIRA

Näheres zur Theorie des Multira-Algorithmus findet sich in Kapitel 2; in Kap. 6 finden sich einige Anwendungsbeispiele.

Sind ein Datensatz und eine Q-Matrix bestimmt worden, kann aus dem Menü **"Programm"** der Befehl mit dem gewünschten Programm ausgewählt werden. Die Wahl des Befehls "MULTIRA" führt zum Fenster *"MULTIRA von C. H. Carstensen und J. Rost"*. Hier findet sich ein Register mit den unterschiedlichen Registerkarten "Daten", "Modell", "Algorithmus", "Bootstrap" und "Ergebnisse".

4.3.1.1 Die Registerkarte "Daten"

Auf der Registerkarte "Daten" wird der gewählte Datensatz mit seinen Eigenschaften aufgeführt. In diesem Fenster ist die Itemauswahl noch veränderbar: In dem entsprechenden Kontrollfeld kann die Liste der Items für die folgende Analyse geändert werden.

ACHTUNG: Werden nicht alle Items ausgewählt, wird eine neue, reduzierte Designmatrix aus den Zeilen der ausgewählten Items aus der Designmatrix Q zusammengestellt.

Ferner kann die Zahl der Personen der analysierten Stichprobe reduziert werden. Im Falle einer Reduktion auf *n* Personen werden Daten der ersten *n* Personen verwendet.

4.3.1.2 Die Registerkarte "Modell"

Auf der Registerkarte "Modell" ist zunächst der Name/Pfad der für den Job ausgewählten Designmatrix Q aufgeführt, die Wahl einer Designmatrix kann auch hier geändert werden.

Die Itemparameter können per CML-Schätzung oder per UML-Schätzung bestimmt werden. Für Daten mit mehrkategoriellem Antwortformat ist wählbar, ob die Schwellenparameter nach einem ordinalen Modell oder einem Äquidistanzmodell geschätzt werden sollen (für dichotome Daten ist diese Auswahl irrelevant).

Zur Schätzung der Personenparameter können weighted Likelihood-Schätzer nach Warm (1989) "wL-Schätzer" oder unbedingte Likelihood-Schätzer "uL-Schätzer" berechnet werden.

Das Kontrollkästchen „**reduzierte Personenzahl analysieren**" ermöglicht es, die Personen mit Extremscores von der Schätzung der Personenparameter auszuschließen (Das ist hauptsächlich zur Kontrolle der Ergebnisse bei einer größeren Zahl von Personen mit Extremscores sinnvoll).

Hinweis: Personen mit Extremscores werden zur Schätzung der Itemparameter von beiden Algorithmen ausgeschlossen. wL-Personenparameter nach Warm (1989) können auch für Personen mit Extremscores geschätzt, uL-Personenparameter dagegen nicht. Um dennoch Schätzungen für Personen mit Extremscores zu erhalten, werden die Scores dieser Personen um den Betrag 0.3 erhöht bzw. reduziert.

Es muß eine der drei Normierungsbedingungen gewählt werden, die die Eindeutigkeit zwischen den Dimensionen und den Itemparametern herstellen. Werden CML-Schätzer angefordert, so muß eine Normierungsbedingung in den Itemparametern gewählt werden: die Summennormierung oder eine Fixierung.

Hinweis: Wird die Fixierung von Items gewählt, muß für jede Dimension (in deren Reihenfolge) ein Item angegeben werden, welches fixiert werden soll. Dieses Item muß in dieser Dimension ein von Null verschiedenes Gewicht haben (Stürzt der Algorithmus während der Schätzung mit einem Fehler ab, war möglicherweise für eine Dimension ein Item mit Nullgewicht angegeben).

4.3.1.3 Die Registerkarte „Algorithmus“

erlaubt es, die maximale Anzahl von Iterationen (1 - 1000) und als Abbruchkriterium füer die Schätzung die maximale Größe der Likelihooddifferenz zwischen zwei zu wählen.

Die weiteren Optionen dienen einer vorläufigen Schrittweitensteuerung des Algorithmus sollen normalerweise nicht verändert werden. Die Schaltfläche "Algorithmus-Standard-Einstellungen" liefert empfohlene Standardeinstellungen.

Hinweis: Endet eine Parameterschätzung nach wenigen Iterationen mit eine negativen Likelihooddifferenz, kann es hilfreich sein, die „Schrittweite zu Beginn“ kleiner zu wählen und die Schätzung noch einmal zu starten (Diese Option sollte nur zusammen mit einer Kontrolle des Iterationsprotokolls verwendet werden. Erfahrungsgemäß wird die Veränderung dieser Einstellungen in vielen Fällen von Modellen mit Konvergenzproblemen nur wenig helfen).

4.3.1.4 Die Registerkarte „Bootstrap“

Das Bootstrapverfahren prüft die Anpassungsgüte eines Modells an die beobachteten Daten. Es erzeugt durch das Resimulieren von Datensätzen und der Schätzung deren Parameter empirisch eine Prüfverteilung für zwei Abweichungsstatistiken zwischen den beobachteten und den unter Modellgeltung zu erwartenden Patternhäufigkeiten. Die Statistiken sind der Pearson χ^2 und eine χ^2-Statistik nach Cressie und Read (siehe von Davier, 1997), bezeichnet mit CR(1) bzw. CR(2/3).

Die Zahl der Resimulationen kann angegeben werden ("Zahl der Bootstrapsamples") und sollte möglichst hoch sein. Das Abbruchkriterium ("Genauigkeit") der Parameterschätzung der resimulierten Daten kann gewählt werden. Um Zeit zu sparen, kann eine größere Likelihood-Differenz zugelassen werden, als bei einer Parameterschätzung mit „Originaldaten“.

Es kann gewählt werden, daß alle Ergebnisse jeder Resimulation in die Ergebnisdatei geschrieben werden ("alle Ergebnisse der Bootstrap-Samples im Output") und es besteht die Möglichkeit, zu jeder Resimulation die Parameter und Datensätze separat zu speichern ("Parameter der Bootstrap-Samples speichern"). Für die letzte Option ist die Zahl der Resimaulaionen auf 99 beschränkt, die Parameter für die Resimulation xx werden in Dateien mit der Endung >.pxx< gespeichert und die resimulierten Datensätze in Dateien mit der Endung >.rxx<.

4.3.1.5 Die Registerkarte „Ergebnisse“

Auf der Registerkarte „Ergebnisse“ läßt sich mittels der entsprechenden Kontrollkästchen einstellen, ob

- das Iterationsprotokoll in der Ergebnisdarstellung unterdrückt werden soll ("kurze Ergebnisdarstellung"),

- die ggf. schon vorhandene Ergebnisdatei überschrieben werden soll,

- Kovarianzen der Itemresiduen (zur Item Fit Kontrolle) ausgegeben werden sollen,

- die Streuungen der Itemparameter ausgegeben werden sollen,

- die Patternhäufigkeiten in eine Datei mit der Endung >.pat< gespeichert werden sollen und ob

- eine Tabellenausgabe in einer Datei namens >Stammname.tab< gespeichert werden soll (Diese Option wird in Zukunft zur Verfügung stehen).

In dem Kontrollfeld „Stammname des Jobs“ besteht die Möglichkeit, den Namen und den (Speicher-) Pfad des Jobs zu ändern. Mit dieser Einstellung ändern sich Name und Pfad für alle Ausgabedateien, als Beispiel wird der Name der Ergebnisdatei angezeigt.

Ein **Job** bezeichnet alle Definitionen zu einem Datensatz inklusive Designmatrix Q, weitere Modellspezifikationen und Programmeinstellungen. Ein Job kann später geladen und/oder ausgeführt werden.

4.3.1.6 Starten der Parameterschätzungen

Auf der rechten Seite des Fensters *"MULTIRA ..."* befinden sich die vier Schaltflächen "Save & Run", „Starten“, „Speichern“ und „Abbrechen“ des Jobs. Der durch die Registerkarten des Fensters beschriebene Job kann gestartet werden und/oder gespeichert werden, um zu einem späteren Zeitpunkt wieder geladen oder ausgeführt zu werden. Beim Speichern erhält man die Möglichkeit, einen Dateinamen für den Stammnamen (s. oben) der Ergebnisdarstellung und einen für die Job-Datei des zu speichernden Jobs auszuwählen.

4.3.2 MKAT

Der MKAT Algorithmus schätzt die Parameter des mehrkategoriellen logistischen Modells von G. Rasch (1961), eine nähere Beschreibung des Modells und des Algorithmus findet sich in Kap. 3). Dazu müssen die gleichen Antwortkategorien für alle Items vorliegen, es dürfen maximal 15 verschiedene Antwortkategorien definiert sein.

Wurde ein Datensatz zur Verfügung gestellt, kann das "Programm" **MKAT** aufgerufen werden. Mit dem Aufruf öffnet sich ein Dialogfenster mit „Registerkarten“, mit welchem der anstehende MKAT-Job definiert und gestartet werden kann.

Mit der Schaltfläche MKAT wird der Job gestartet; die Möglichkeit den Job zu speichern, wird in Zukunft zur Verfügung stehen.

4.3.2.1 Die Registerkarte "Daten"

Auf der Registerkarte "Daten" wird der gewählte Datensatz mit seinen Eigenschaften aufgeführt. In diesem Fenster ist die Itemauswahl noch veränderbar: In dem entsprechenden Kontrollfeld kann die Liste der Items für die folgende Analyse geändert werden.

Ferner kann die Zahl der Personen der analysierten Stichprobe reduziert werden. Im Falle einer Reduktion auf *n* Personen werden Daten der ersten *n* Personen verwendet.

4.3.2.1 Die Registerkarte "Algorithmus"

Hier kann die maximale Anzahl von Iterationen (1 - 1000) festgelegt werden und die maximale Größe der Likelihooddifferenz zwischen zwei Iterationen als Abbruchkriterium füer die Schätzung gewählt werden.

4.3.2.1 Die Registerkarte "Ergebnisse"

Es kann gewählt werden und ob die Ergebnisdatei durch die anstehende Ergebnisausgabe überschrieben werden soll.

Der Stammname des Jobs kann editiert werden, um den Namen und den (Speicher-) Pfad des Jobs zu ändern. Das betrifft die Ergebnisdatei und die Datei, in die die Personenparameter geschrieben werden.

4.3.3 MLT

Der Martin Loef Test überprüft die Modellannahme der Itemhomogenität des Raschmodells für dichotome Daten anhand einer Aufteilung in zwei Untertests. Näheres dazu in Kapitel 3 (S.22). In MULTIRA steht kein eigenes Unterprogramm für den MLT zur Verfügung. Stattdessen kann ein generalisierter MLT durchgeführt werden, indem ein bedingt geschätztes eindimensionales Rasch-Modell gegen ein bedingt geschätztes zwei- oder mehrdimensionales Rasch-Modell mit einem LR Test getestet wird. Das mehrdimensionale Modell muß jedes Item des eindimensionalen Modells genau einer latenten Dimension zuordnen.

Weitere Details zum generalisierten MLT geben Carstensen (2000) und Carstensen und Rost (in Vorbereitung).

4.3.4 LLTM

Der Algorithmus zum Linear Logistischen Testmodell für dichotome Daten von G. Fischer findet sich im Lehrbuch (1974). Nicht alle Optionen wurden in das Menü aufgenommen.

Wurden ein Datensatz zur Verfügung gestellt und eine Designmatrixdatei geöffnet, kann das „Programm“ LLTM aufgerufen werden. Es öffnet sich ein Dialog-Fenster mit vier Registerkarten (ähnlich wie im Dialog-Fenster zum Programm MULTIRA).

Die **Registerkarte „Daten“** bietet die Möglichkeit, Items für die folgende Parameterschätzung auszuwählen.

Die **Registerkarte „Modell“** beinhaltet den Namen der Designmatrixdatei; diese muß (trotz anderer Bedeutung) das gleiche Format haben, wie für den MULTIRA-Algorithmus. Die Option „Einheitsmatrix als Designmatrix“ (entspricht der Einstellung EMATRIX=1 und) bewirkt, das ein einfaches Rasch-Modell geschätzt wird.

Auf der **Registerkarte „Algorithmus“** lassen sich die maximale Anzahl von Iterationen und die zu erreichende Genauigkeit als Abbruchkriterien angeben.

Die **Registerkarte „Ergebnisse“** bietet die Möglichkeit, den Namen der Ergebnisdatei zu verändern und auszuwählen, ob die Ergebnisdatei mit der neuen Parameterschätzung überschrieben werden soll.

Mit den Schaltflächen am rechten Fensterrand läßt sich die Parameterschätzung nach Auswahl der Optionen starten oder abbrechen.

4.3.5 Simulator

Der Simulator generiert Daten nach dem MULTIRA-Modell. Nachdem die Personen- und Itemparameter festgelegt wurden, werden nach der Modellgleichung mit einem Zufallszahlengenerator die Antwortdaten erzeugt.

Wurde eine Designmatrix für das Modell gewählt, nach dem die Daten erzeugt werden sollen, kann das Programm „Simulator“ aufgerufen werden. Zunächst muß ein Dateiname für die zugenerierenden Daten bestimmt werden. In einer Datei mit gleichem Namen und der Endung >.sim< werden die zugrundeliegenden Parameter gespeichert und können später mit den geschätzten Parametern verglichen werden.

Die Personenparameter können nach einer Normalverteilungsannahme zufällig generiert werden, wobei Mittelwert und Streuung einstellbar sind. Die Parameter können auch aus einer Datei eingelesen werden. Die Datei muß im Textformat gespeichert sein und die Parameterwerte müssen nach einer Zeile, deren erste sechs Zeichen "person" in Groß- oder Kleinschreibung enthalten, achtspaltig abgespeichert sein mit je einer Zeile je Person

Die Itemparameter können nach einem Muster, welches mit den Optionen "Streubreite der Itemparameter" und "Abstand der Schwellen" variiert werden kann generiert werden. Sie können ebenfalls aus einer Datei eingelesen werden. Dazu müssen sie in derselben Textdatei wie für die Personenparameter nach einer Zeile, deren erste vier Zeichen "Item" oder "Schw" in Groß- oder Kleinschreibung enthalten, achtspaltig abgespeichert sein.

Wird eine Parameterdatei verwendet, kann diese Itemparameter enthalten, sie muß in jedem Fall Personenparameter enthalten.

4.4 Das Menü „Fenster“

Mit den Befehlen **Daten** und **Ergebnisse** lassen sich die analysierten Daten und die aktuelle Ergebnisdatei in einem Textfenster darstellen. Der Inhalt dieser Textfenster kann nicht verändert werden. Soll eine Daten- oder Ergebnisdatei verändert werden, muß sie in einem Editor-Fenster geöffnet werden.

Der Befehl **Anordnen** ordnet alle geöffneten Fenster kaskadenförmig an.

4.5 Das Menü „?“

Das Menü „**?**“ beinhaltet die Befehle **Hilfe** und **Über MULTIRA.**

Hilfe

Der Befehl **Hilfe** öffnet dieses Handbuch als Acrobat PDF-Datei.

Über MULTIRA

Hier finden sich Informationen über MULTIRA, insbesonder die Versionsnummer und das Datum der Kompilierung. Außerdem sind die Kontaktadressen der Autoren angegeben, da Fragen und Anregungen zu MULTIRA sind jederzeit willkommen sind.

5. Darstellung einer Ergebnisdatei

Im folgenden wird eine Ergebnisdatei dargestellt und erläutert, die sich nach einer Analyse des Beispieldatensatzes >beispiel.dat< mit der zugehörigen Designmatrix Q in >beispiel.qmx< ergeben kann. Die Itemparameter wurden mit dem UML Algorithmus geschätzt und für die Personenparameter wurden wL-Schätzer berechnet.

```
//   MULTIRA v1.45 - Ein Programmsystem für mehrdimensionale Testmodelle
//
//  geschrieben von Claus H. Carstensen und Jürgen Rost
//  (c) 1995 -1999 IPN Kiel, kompiliert: 10.12.00
//
//  Das MULTIRA-Programm beinhaltet folgende Algorithmen:
//
//  MULTIRA (C. H. Carstensen, J. Rost) zum MULTIRA-Modell von Jürgen Rost (1996)
//  MKAT (J. Rost) zum Mehrdimensionalen-Rasch-Modell von G.Rasch (1961)
//  Martin Löf Test für Itemhomogenität
```

Der Inhalt der Ergebnisdatei ist in Courierschrift dargestellt, die Erläuterungen in Arialschrift. Nach den Kopfzeilen werden das Datum der Analyse, Eigenschaften des Datensatzes und das gerechnete "Programm" wiedergegeben.

```
Datenanalyse vom 11/12/00, um 17:00:46 Uhr

Daten      : C:\multira\beispiel.dat
Ergebnisse : C:\multira\beispiel.out
Anzahl von Personen   :   500
Anzahl von Kategorien :     4
Anzahl von Items      :    15
Items in Spalten:1 3 5 7 9 11 13 15 17 19 21 23 25 27 29

M U L T I R A - Algorithmus, C. H. Carstensen und J. Rost
```

Für die folgend dargestellte Analyse wurde der MultiRa-Algorithmus verwendet.

```
verwende Q-Matrix in C:\multira\beispiel.qmx

 Ein Beispiel für eine Design-Matrix Q für sample.dat,
 einem simulierten Datensatz nach den Parametern in sample.si

 Nach der Abschnittsmarkierung folgt unmittelbar die Q-Matrix
 nach der Q-Matrix soll eine leere Abschnittsmarkierung [] st

 Die Koeffizienten werden durch ein Leerzeichen oder ; getren
 Dezimalstellen durch . oder , abgetrennt.

  1.00  0.00
  1.00  0.00
  1.00  0.00
  1.00  0.00
  1.00  0.00
  0.00  1.00
  0.00  1.00
  0.00  1.00
  0.00  1.00
```

```
  0.00  1.00
  1.00  1.00
  1.00  1.00
  1.00  1.00
  1.00  1.00
  1.00  1.00

Die Q-Matrix definiert    2 Komponente/n für   15 Items
```

Es wird der Kommentartext aus der verwendeten Designmatrix-Datei wiedergegeben und die eingelesene Matrix Q dargestellt.

```
Korrelationen der Komponenten untereinander
Komp.     1
    2 -0.50
```

Diese Angabe ist unwesentlich.

Es folgen Angaben über das gewählte Modell und über Anzahlen von Personen, der zu schätzenden Modellparameter, aufgeschlüsselt nach Item- und Personen/Score- parametern und der unterschiedlichen Antwortmuster.

```
nachfolgend geschätztes Modell:
Itemparameter geschätzt durch        UML
Schwellenparameter geschätzt nach dem Ordinal-Modell
Personenparameter geschätzt durch    wL-Schätzer
Itemparameter-Restriktionen:         Items summennormiert

Anzahl der Personen im Datensatz                    :   500
... die ausschliesslich Extremscores erzielt haben  :    20
... in der folgenden Parameterschätzung             :   480

Anzahl der (restringierten) Dimensionen          :        2
Anzahl der unabhängigen Itemparameter            :       43

Anzahl der beob. Scorevektoren ohne Extremscores :      322
Anzahl der beobachteten Scorevektoren            :      337      =960+43
Anzahl der unabhängigen Scorevektoren            :      960

Anzahl der geschätzten Modellparameter CML       :       43      =2*480+43
Anzahl der geschätzten Modellparameter MML       :     1003
Anzahl der geschätzten Modellparameter UML       :     1003
Anzahl der geschätzten Modellparameter WML       :     1043
                                                                 =2*500+43
Anzahl verschiedener beobachteter Antwortmuster  :      499
maximale Zahl verschiedener Antwortmuster      :  1073741824
```

In der folgenden Tabelle werden die Kategorienhäufigkeiten wiedergegeben.

```
### Kategorienhäufigkeiten ###

Item  1:    60  114  170  136   E:   1.80
Item  2:    59  138  165  118   E:   1.71
Item  3:    68  141  143  128   E:   1.69
Item  4:    71  129  168  112   E:   1.67
Item  5:    82  146  139  113   E:   1.59
Item  6:    78  155  143  104   E:   1.57
Item  7:    92  142  150   96   E:   1.52
Item  8:   103  156  123   98   E:   1.45
```

```
Item   9:    97  150  155   78   E:   1.45
Item  10:   119  152  122   87   E:   1.37
Item  11:   134  142   95  109   E:   1.37
Item  12:   145  131  106   98   E:   1.33
Item  13:   148  136  115   81   E:   1.27
Item  14:   156  125  107   92   E:   1.28
Item  15:   164  122  103   91   E:   1.25
```

Wie an den Häufigkeiten (und auch am Erwartungswert in der letzten Spalte) zu erkennen ist, ist das erste Item das leichteste und das letzte Item das schwierigste.

Das folgende Iterationsprotokoll läßt mit der Option "kurze Ergebnisdarstellung" in der Registerkarte "Ergebnis" unterdrücken. Ausgegeben werden die Loglikelihood und der Loglikelihoodunterschied von Iteration zu Iteration ("`-diff.`").

```
gemeinsame uL-Maximierung für Items und   480 Personen

Iteration      log-uL        -diff.      step        SuGes        SuDiv      SuItems
      1     -8078.8462 ************   0.5000 287.6983032 287.6983032  10.0765276 0.13556 0.24472
      2     -7477.9105  600.9357299   0.5000 169.6179810 166.6086426   8.9165173 0.09966 0.23182
      3     -7126.6603  351.2502402   0.5000 138.4605865 133.7366791   7.6382604 0.11197 0.24400
    ...
     54     -6400.9052    0.0001503   1.0000   0.0728525   0.0708986   0.0082558 0.00012 0.00009
     55     -6400.9051    0.0001155   1.0000   0.0637906   0.0620803   0.0072273 0.00011 0.00008
     56     -6400.9050    0.0000888   1.0000   0.0558519   0.0543507   0.0063258 0.00009 0.00007

   56 Iterationen benötigt, um eine loglike-Differenz von    0.0000888 zu erreichen
```

Die zweite und dritte Spalte geben logarithmierte Likelihood und die Differenz der log. L. zwischen der letzten und vorletzten Iteration wieder. Die weiteren Spalten sind für den Anwender nicht weiter von Bedeutung („Step“ gibt die Schrittweite je Iteration an, sie wird vom Algorithmus nach den Kriterien in der Registerkarte „Algorithmus“ verändert).

```
Itemparameter (k-1)/k korrigiert
```

Nach einer UML-Schätzung der Itemparameter werden diese um den Faktor (k-1)/k korrigiert. Eine wL-Schätzung der Personenparameter ergibt das folgende Iterationsprotokoll. Sowohl wL-Schätzer als auch uL-Schätzer können in MULTIRA auch für Personen mit Minimal- oder Maximalscores geschätzt werden (siehe S.31). Die folgenden Statistiken werden für die mit dieser Option gewählte Zahl von Personen berechnet.

```
wL Schätzung der Personenparameter für   500 Personen
      1    -10538.0704 ************   0.2500 305.8689270 305.8689270   0.0000000 0.00009 0.00007
      2     -9390.3918 1147.6785613   0.2500 210.1355438 209.5812683   0.0000000 0.00009 0.00007
      3     -8743.0480  647.3437991   0.2500 157.0072327 154.9534607   0.0000000 0.00009 0.00007
    ...
     66     -7309.4057    0.0001248   1.0000   0.0009122   0.0009030   0.0000000 0.00009 0.00007
     67     -7309.4056    0.0001063   1.0000   0.0007818   0.0007766   0.0000000 0.00009 0.00007
     68     -7309.4055    0.0000913   1.0000   0.0006707   0.0006615   0.0000000 0.00009 0.00007

   68 Iterationen benötigt, um eine loglike-Differenz von    0.0000913 zu erreichen
```

```
Die folgenden Angaben beziehen sich auf   500 Personen der Stichprobe

Item- und Personenparameter (WL-Schätzer) geschrieben in C:\multira\beispiel.par
```

Die geschätzten Parameter und die gewichteten Scores werden in einer Text-Datei >... .par< gespeichert. Die folgende Tabelle zeigt Mittelwerte und Streuungen sowie die Reliabilität für jede Dimension. Die Reliabilitäten werden anhand der Maximum-Likelihood-Schätzfehler der Itemparameter bestimmt.

```
### Mittelwerte, Streuungen & Reliabilität für die Personenparameter ###
Komp.      M      SD     Rel    Mittelwert der mittleren Itemparameter in j
   1   0.099   1.713   0.887   0.000
   2  -0.295   1.734   0.895   0.000
```

Anschließend werden Korrelationen zwischen den Personenparametern einmal wie beobachtet und zweitens anhand der Reliabilitätsschätzungen minderungskorrigiert angegeben. Außerdem werden die Korrelationen zwischen den gewichtenen Scores und zwischen Scores und Parametern je Dimension dargestellt.

```
### Korrelationen der Personenparameter für   2 Komponenten
Komp.     1
    2  0.00


### Minderungskorrigierte Korrelationen der Personenparameter
Komp.     1
    2  0.00


### Korrelationen der gewichteten Personenscores
Komp.     1
    2  0.77


### Korrelationen der gewichteten Scores mit den Personenparametern
Komp.
    1   0.882
    2   0.898
```

Die folgende Tabelle gibt die Schwellenparameter (dekumulierte Itemparameter) wieder. Der mittlere Itemparameter ergibt sich aus der Summe der Schwellenparameter dividiert durch deren Anzahl und kann als schwellenunabhängige Itemschwierigkeit aufgefasst werden.

```
### Schwellenparameter und mittlerer Itemparameter ###
    -> gemäß dem ordinalen Modell

Item   1:   -2.000  -0.614   1.143  M:  -0.490
Item   2:   -2.131  -0.296   1.365  M:  -0.354
Item   3:   -1.940  -0.100   1.144  M:  -0.299
Item   4:   -1.815  -0.322   1.466  M:  -0.224
Item   5:   -1.668   0.085   1.340  M:  -0.081
Item   6:   -2.344  -0.397   1.175  M:  -0.522
Item   7:   -2.038  -0.465   1.346  M:  -0.386
```

```
Item   8:   -1.914  -0.075   1.201  M:  -0.263
Item   9:   -1.979  -0.336   1.710  M:  -0.202
Item  10:   -1.653   0.021   1.406  M:  -0.075
Item  11:   -1.690   0.416   1.506  M:   0.077
Item  12:   -1.464   0.313   1.801  M:   0.217
Item  13:   -1.427   0.390   2.229  M:   0.398
Item  14:   -1.261   0.346   1.942  M:   0.342
Item  15:   -1.123   0.417   1.947  M:   0.414

 Itemparameter summennormiert in  2 Dimension/en
```

In der letzten Spalte sind Mittelwerte der Schwellenparameter angegeben: sie können als Itemschwierigkeitsparameter verwendet werden.

Ferner wird die verwendete Normierungsbedingung zwischen den Items und den latenten Dimensionen wiedergegeben. (wurde die Fixierung der Mittelwerte der Personenparameter je Dimension gewählt, so kann man das auch in der Tabelle mit Mittelwerten für die Personenparameter weiter oben erkennen).

Die nächste Tabelle gibt die Itemresiduenstatistiken "Outfit" und "Intfit" wieder. Beide Statistiken sagen etwas über die Trennschärfen eines Items aus (siehe S. 18). Der Outfit, eine (ungewichtete) Statistik der Residuen eines Items und aller Personen mit Mittelwert Eins, wird in der ersten Spalte wiedergegeben, seine z-Transformierte (M=0, SD=1) in der zweiten Spalte. Der Intfit, eine gewichtete Statistik der Residuen eines Items und aller Personen (M=1), wird in der dritten Spalte wiedergegeben und ein Standardwert (M=0, SD=1) dazu in der vierten Spalte. Durch die Gewichtung mit der Antwortwahrscheinlichkeit im Infit wird der Einfluß der Residuen von Personen mit unwahrscheinlicheren, d.h. extremen Antworten Parametern gegenüber dem Outfit abgewertet.

```
### Item Residuen Fit Statistik ###

    -> anhand der ML Personenparameter
          unweighted            weighted
Item   Outfit    t(i)       Infit    t(i)
   1   0.950   -0.564       0.913   -1.391
   2   0.986   -0.145       1.022    0.374
   3   0.939   -0.691       0.940   -0.959
   4   1.036    0.457       1.012    0.209
   5   0.971   -0.320       0.972   -0.429
   6   1.052    0.641       1.049    0.786
   7   0.973   -0.304       0.928   -1.144
   8   0.983   -0.166       1.016    0.265
   9   0.991   -0.088       0.982   -0.260
  10   0.858   -1.700       0.892   -1.725
  11   1.072    0.649       1.000    0.017
  12   0.963   -0.282       1.025    0.383
  13   0.920   -0.686       0.948   -0.751
  14   1.094    0.789       1.007    0.125
  15   0.992   -0.020       1.015    0.236
```

```
    -> anhand der WL Personenparameter
            unweighted              weighted
Item    Outfit    t(i)        Infit    t(i)
   1    0.898   -1.261        0.886   -1.865
   2    0.943   -0.700        0.986   -0.214
   3    0.889   -1.382        0.910   -1.485
   4    0.977   -0.260        0.978   -0.343
   5    0.917   -1.024        0.937   -1.027
   6    0.991   -0.084        1.006    0.123
   7    0.908   -1.169        0.897   -1.699
   8    0.928   -0.872        0.973   -0.416
   9    0.935   -0.830        0.946   -0.868
  10    0.819   -2.340 -!     0.858   -2.353 -!
  11    0.952   -0.461        0.951   -0.718
  12    0.888   -1.139        0.971   -0.406
  13    0.843   -1.677        0.903   -1.492
  14    0.969   -0.269        0.956   -0.629
  15    0.907   -0.859        0.961   -0.548
```

Dit Itemfitmaße werden anhand der ul-Schätzer für die Personeparameter und/oder nach den wL-Schätzern für die Personenparameter berechnet, je nachdem, welche Parameter vorher geschätzt wurden. Bei den mit einem Ausrufezeichen gekennzeichneten Items weichen die Residuen signifikant (5%, ohne Korrektur für multiples Testen) von ihrem Erwartungswert ab. Sind die Werte größer als Eins bzw. Die transformierten Fitstatistken positiv, so ist das betreffende Item zu trennschwach.

Da die uL maximiert wurde, können keine Likelihoodvergleiche zur Modellgeltungskontrolle durchgeführt werden. Stattdessen kann das **Bootstrap**-Verfahren (siehe S.18) eingesetzt werden. Dazu werden zwei χ^2-Statistiken ("CR(1)" und "CR(2/3)") verwendet, die die Unterschiede zwischen den unter dem Modell erwarteten Anworten und den beobachteten Antworthäufigkeiten aggregieren. Da die asymptotische χ^2-Verteiltheit dieser Prügrößen nicht angenommen werden kann, da bei der sehr großen Zahl von theoretisch beobachtbaren Antwortmustern die erwarteten Häufigkeiten für jedes Antwortmuster zu klein d.h. deutlich kleiner als 5 oder 1 (die üblichen Daumenregeln für χ^{2-}Tests) sind. Die Prüfverteilungen werden durch Resimulationen von Datensätzen und die Schätzung der Parameter inklusive Fitmaße empirisch approximiert. Die folgende Darstellung gibt die Angabe der beiden Fitstatistiken wieder:

```
### Kennwerte zur Modellgeltungskontrolle ###

      log uL          np    geom. Mittel
  -6400.9050        1003          0.4111

  log uL|wLE          np    geom. Mittel
  -6656.8509        1016          0.3967
```

Anschließend sind die Ergebnisse eines Bootstrapverfahren dargestellt:

```
Die Parameter für die folgenden Simulationen C:\multira\beispiel.ref
wurden gelesen aus C:\multira\beispiel.par
```

... (es folgen Meldungen aus den Schätzungen der Bootstrap-Stichproben)

```
In 1 Iterationen war die LogL-Diff. größer als die vorherige (nichtmonotone Konver-
genz)
...

### Ergebnisse des Bootstrapverfahrens ###

                  GoF CR(1)     CR(2/3)          df       np
echte Daten:      1.706E+09   7.446E+06  1073740820     1003
Sample    1:    787.909E+06   5.337E+06  1073740820     1003
Sample    2:      1.353E+09   5.959E+06  1073740826      997
Sample    3:     27.148E+09  21.360E+06  1073740830      993
```

... (Ergebnisse der Bootstrap-Stichproben)

```
Sample   48:      1.202E+09   5.963E+06  1073740826      997
Sample   49:      1.063E+09   5.275E+06  1073740832      991
Sample   50:    589.617E+06   4.547E+06  1073740844      979

Rang(echte Daten) :       42          45  je von   51
Wahrscheinlichkeit:  p=0.18      p=0.12
für einen gleichen oder besseren Modell-Fit der echten Daten unter den resimulier-
ten Datensätzen

RMSE über alle geschätzten Parameter der resimulierten Datensätze:
Personen:    0.6873, Items:    0.2492

MULTIRA-Algorithmus beendet (um 17:21:35 Uhr)
```

Hier wurden 50 Datensätzen resimuliert (für Bootstrapverfahren sollten nach Möglichkeit mehr Datensätze resimuliert werden). Angegeben ist der Rang der Fitstatistiken der echten Daten unter denen der resimulierten Daten. Dieser Rang ist dann noch einmal in die Wahrscheinlichkeit (geschätzt durch die relative Häufigkeit) p umgerechnet angegeben, mit der Fitstatistiken von resimulierten Datensätzen einen „schlechteren" Fit anzeigen, als die die Fitstatistik für die Originaldaten.

Die Wahrscheinlichkeit „schlechter" passender resimulierter Datensätze ist p=0.18 nach CR(1) und p=0.12 nach CR(2/3); sie ist höher als die häufig gewählte Ablehnungsgrenze von 0.05. Das Bootstrapverfahren bestätigt also die Modellanpassung des Datenbeipiels mit der verwendeten Designmatrix - was bei einem modellkonform simulierten Datensatz auch zu erwarten war.

.6. Anwendung in einigen vertiefenden Beispielen

In diesem Umgang mit MULTIRA vertieft werden. Anhand von Beispielen wird auf unterschiedliche mögliche Fragestellungen eingegangen. Die folgende Tabelle gibt eine Übersicht über die Datenbeispiele, die Fragestellungen, die verwendeten Algorithmen und Modellgeltungstests.

Datensatz	Kap.	Fragestellung	Verfahren
Phyikleistungstest, 20 Items, n=108,	6.1.1	Skalierung	UML Schätzung
	6.1.2	Skalierung	CML Schätzung
	6.1.3	Dimensionalität	Informationskriterien
	6.1.4	Dimensionalität	LR - Tests
?-fragebogen	6.2	Itemhomogenität	CML Schätzung und LR-Test, ein generaliserter MLT
Intelligenztest BIS 45 Aufgaben in 7 Dimensionen und 3x4 Facetten	6.3	Skalierung, emp. Bestätigung der Modellannahmen	UML, Bootstrap

Das erste Datenbeispiel ist ein Leistungstest, der aus zwanzig Physikaufgaben besteht und von einer relativ kleinen Stichprobe bearbeitet wurde. Der Test ist mehrdimensional konzipiert, jedoch zeigen unterschiedliche Modellgeltungstests, daß der Test in der untersuchten Stichprobe nur eine latente Dimension anspricht. Die Itemparameter in diesem Datenbeispiel können mit beiden Algorithmen, CML und UML, geschätzt werden. Zur Überprüfung der Dimensionalität dieses Tests werden Likelihood ratio Tests, Informationskriterien und daws Bootstrapverfahren verwendet.

6.1 Eine, zwei oder vier Dimensionen ? Ein Physikleistungstest

- wird ergänzt –

6.2 Ein Physikinteressenfragebogen

- wird ergänzt –

6.3 Der Test zum Berliner Intelligenzstrukturtest - Ein Test im Facettendesign

- wird ergänzt -

Literatur

Andrich, D., 1978. A rating formulation for ordered response categories. *Psychometrika* 43, S. 561-573.

Andrich, D., 1982. An extension of the Rasch Model for ratings providing both location and dispersion parameters. *Psychometrika* 47, S. 105-113.

Carstensen, Claus H., (2000). *Mehrdimensionale Testmodelle mit Anwendungen aus der pädagogisch-psychologischen Diagnostik*. Kiel: IPN, Schriftenreihe 171.

Davier, M. v., 1997. Bootstrapping Goodness-of-Fit Statistics for Sparse Categorical Data - Results of a Monte Carlo Study. *Methods of Psychological Research - MPR online*, Vol. 2, No. 2, PABST SCIENCE PUBLISHERS (http://www.pabst-publishers.de/mpr/index.html).

Fischer, G. H. (1972). A measurement model for the effect of mass-media. *Acta Psychologica*, 36, 207-220.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche (2nd Edition, Chicago, University of Chicago Press, 1980).

Rasch, G. (1961). *On general laws and the meaning of measurement in psychology*. Berkeley: University of California Press.

Rost, J. (1996a). *Lehrbuch Testkonstruktion Testtheorie*. Bern: Huber.

Rost, J. & Carstensen, C. H., (in press). Multidimensional Rasch Measurement via Item ComponentModels and Faceted Designs. *Applied Psychological Measurement*.

Stegelmann, W. (1983). Expanding the Rasch model to a general model having more than one dimension. *Psychometrika*, 48, 2, 259-267.

Verhelst, N D., Glas, C. A W., Verstralen, H. H. F. M., 1994. *OPLM*: Computer program and manual. Arnhem, CITO.

Warm, T. A. (1989). Weighted Likelihood Estimation of Ability in Item Response Theory. *Psychometrika*, 54, 3, 427-450.

Wright, B. D., Masters, G. N. (1982). *Rating Scale Analysis Rasch Measurement*. Chicago, MESA Press.

Stichwort-Register