# Das allgemeine ordinale Rasch-Modell

P14.8

Kurt Holm

Almo Statistik-System
www.almo-statistik.de
holm@almo-statistik.de
kurt.holm@jku.at

**2014**

Siehe auch das Almo-Dokument 7a. "Wie man mit Almo ein Rasch-Modell rechnet"
Im Text wird häufig auf das Dokument **P0** Bezug genommen. Dabei handelt es sich um das Almo-Dokument "Arbeiten mit Almo.PDF" (Dokument 0).

Siehe insbesondere auch das Almo-Dokumente 20 **Latent structure analyssis.PDF** (Hermann Denz), in dem das zum Rasch-Modell sehr ähnliche "Ogiven-Modell" dargestellt, sowie das Almo-Dokument 19 **Guttman- und Mokken-Skalierung.PDF** (Joachim Gerich

## Weitere Almo-Dokumente

Die folgenden Dokumente können alle von der Handbuchseite in http://www.almo-statistik.de/ heruntergeladen werden

```
0.  Arbeiten_mit_Almo.PDF (1 MB)
1a. Eindimensionale Tabellierung.PDF  (1.8 MB)
1b. Zwei- und drei-dimensionale Tabellierung.PDF (1.1 MB)
2.  Beliebig-dimensionale Tabellierung.PDF (1.7 MB)
3.  Nicht-parametrische Verfahren.PDF (0.9 MB)
4.  Kanonische Analysen.PDF (1.8 MB)
    Diskriminanzanalyse.PDF (1.8 MB)
    enthält: Kanonische Korrelation, Diskriminanzanalyse, bivariate
    Korrespondenzanalyse, optimale Skalierung
5.  Korrelation.PDF (1.4 MB)
6.  Allgemeine multiple Korrespondenzanalyse.PDF (1.5 MB)
7.  Allgemeines ordinales Rasch-Modell.PDF (0.6 MB)
7a. Wie man mit Almo ein Rasch-Modell rechnet.PDF (0.2 MB)
8.  Tests auf Mittelwertsdifferenz,  t-Test.PDF (1,6 MB)
9.  Logitanalyse.pdf  (1,2MB)  enthält Logit- und Probitanalyse
10. Koeffizienten der Logitanalyse.PDF  (0,06 MB)
11. Daten-Fusion.PDF  (1,1 MB)
12. Daten-Imputation.PDF  (1,3 MB)
13. ALM Allgemeines Lineares Modell.PDF  (2.3 MB)
13a. ALM Allgemeines Lineares Modell II.PDF  (2.7 MB)
14. Ereignisanalyse: Sterbetafel-Methode, Kaplan-Meier-Schätzer,
    Cox-Regression.PDF (1,5 MB)
15. Faktorenanalyse.PDF (1,6 MB)
16. Konfirmatorische Faktorenanalyse.PDF (0,3 MB)
17. Clusteranalyse.PDF (3 MB)
18. Pisa 2012 Almo-Daten und Analyse-Programme.PDF (17 KB)
19. Guttman- und Mokken-Skalierung.PFD (0.8 MB)
20. Latent Structure Analysis.PDF (1 MB)
21. Statistische Algorithmen in C (80 KB)
22. Conjoint-Analyse (PDF 0,8 MB)
23. Ausreisser entdecken (PDF 170 KB)
24. Statistische Datenanalyse Teil I, Data Mining I
25. Statistische Datenanalyse Teil II, Data Mining II
26. Statistische Datenanalyse Teil III, Arbeiten mit Almo-
    Datenanalyse-System
27. Mehrfachantworten, Tabellierung von Fragen mit Mehrfach-
    antworten (0.8 MB)
28. Metrische multidimensionale Skalierung (MDS)  (0,4 MB)
29. Metrisches multidimensionales Unfolding (MDU) (0,6 MB)
30. Nicht-metrische multidimensionale Skalierung (MDS)  (0,5 MB)
31. Pfadanalyse  (0,7 MB)
32. Datei-Operationen mit Almo  (1,1 MB)
33. Wählerstromanalyse und Wahlhochrechnung (1,6 MB)
34. Soziometrie. Auswertung soziometrischer Daten (0,5 MB)
```

# Inhaltsverzeichnis

**Kurt Holm**

**P14.8 Das allgemeine ordinale Rasch-Modell**
P14.8.0 Programm und Programm-Autoren 4
P14.8.1 Items für das ordinale Rasch-Modell 5
P14.8.2 Almo-Programm-Masken 6
P14.8.3 Die logistische Itemfunktion im Rasch-Modell 7
P14.8.3.1 Die logistische Funktion im dichotomen Rasch-Modell 8
P14.8.3.2 Die Kategorienfunktion im ordinalen Rasch-Modell 11
P14.8.4 Schätzung der Parameter 15
P14.8.5 Vergleich von Almo, Conquest, Multira, Winmira 16
P14.8.6 Eingabe in die Programm-Maske Prog14m3 18
P14.8.7 Eingabe in die Programm-Maske Prog14m4 21
P14.8.7.1 Variablennamen 21
P14.8.7.2 Ein- und Ausschliessen von Untersuchungseinheiten 22
P14.8.7.3 Umkodierungen 23
P14.8.7.4 Behandlung fehlender Werte 24
P14.8.7.5 Modell-Test 26
P14.8.7.6 Itemfit 27
P14.8.7.7 Test auf Personen-Homogenität 33
P14.8.7.7.1 Andersen-Test 35
P14.8.7.7.2 Andersen-Test mit Gruppierungsvariable 40
P14.8.7.7.3 Wald-Test auf Personen-Homogenität 43
P14.8.7.7.4 Martin-Löf-Test auf Item-Homogenität 44
P14.8.7.8 Antwortmuster ausgeben 44
P14.8.7.9 Skalenwerte (Fähigkeitswerte) in Datei geben 46
P14.8.7.10 Rasch-konforme oder Rasch-perfekte Daten erzeugen 50
P14.8.7.11 R-Programm erzeugen 55
P14.8.7.12 Grafik-Optionen 57
P14.8.8 Ausgabe aus Prog14m4 58
P14.8.8.1 Basis-Statistiken 59
P14.8.8.2 Schwierigkeits-, Schwellen- und Fähigkeitsparameter 59
P14.8.8.3 Grafiken bei dichotomen und polytomen Items 64
P14.8.8.4 Itemfit-Masse 71
P14.8.8.4.1 Infit und Outfit 71
P14.8.8.4.2 Q-Index (nach Rost/von Davier) 72
P14.8.8.4.2.1 Die Steigung der logistischen Rasch-Funktion 76
P14.8.8.5 Modelltest und Bootstrap 77

**Heinrich Potuschak**

**P14.8.9 Parameterschätzung und Modelltests 79**
P14.8.9.1 Kodierung der Pattern 79
P14.8.9.2 Schätzung der bedingten Schwierigkeitsparameter 79
P14.8.9.3 Schätzung der gewichteten Personenparameter 81
P14.8.9.4 Standardfehler der Schwierigkeitsparameter 82
P14.8.9.5 Andrichs Reliabilität 82
P14.8.9.6 Modell-Tests 82
P14.8.9.6.1 Globale Modell-Tests 82
P14.8.9.6.2 Itemfit-Tests nach Wright & Masters 83

P14.8.9.6.3 Der Q-Index (nach Rost/von Davier) 84
P14.8.9.6.4 Personenhomogenität: Andersen-Test, Wald-Test 85
P14.8.9.6.5 Itemhomogenität: Martin-Löf-Test 85
P14.8.9.7 Bootstrappen 85
P14.8.9.8 Simulation von Rasch-konformen Daten 86

**Kurt Holm**
**Exkurs: Zentrale Konzepte des Rasch-Modells 87**

**Kurt Holm**
**Exkurs: Vergleich von Rasch-Modell mit "Latent Structure Analysis" 91**

**Literatur zum ordinalen Rasch-Modell 99**

**Kurt Holm**

## P14.8 Das allgemeine ordinale Rasch-Modell

### P14.8.0 Programm und Programm-Autoren

Die für ein ordinales Rasch-Modell zentralen Funktionen für die Ermittlung der Itemparameter, der Personenparameter, der Modell- und Itemfit-Tests und der Datensimulation wurden von **Heinrich Potuschak** in der Programmiersprache „R“ geschrieben. Kurt Holm hat diese Funktionen in die Programmiersprache „C“ übertragen und in das Almo-Rasch-Programm eingebunden. Dieses steht den Benutzern nunmehr mit vielen Optionen und einer detaillierten Ergebnisausgabe mit vielen Grafiken frei zur Verfügung.

Der R-Quelltext von Heinrich Potuschak zusammen mit einer Erläuterung der Funktionen ist im Almo-Unterordner „**Rasch_in_R**“ enthalten. Im letzten Abschnitt P14.8.9 des hier vorliegenden Handbuchs hat Potuschak den Kalkül des ordinalen Rasch-Modells detailliert dargestellt.

Der gesamte C-Quelltext (etwas gekürzt, ohne komplexe Optionen und ohne Grafikteil) wird ebenfalls frei zur Verfügung gestellt. Er ist im Almo-Unterordner **Algorithmen_in_C** in den Modulen "a_up_algorith11.c" und "a_Main_Konsole_24.c" enthalten. Im Modul "a_Main_Konsole_25.c" ist ein Programm zur Simulation von Rasch-konformen Daten enthalten. Alle C- und R-Unterprogramme können von interessierten Programmierern beliebig verwendet werden.

Das Rasch-Programm und der Quelltext kann in www.almo-statistik.de heruntergeladen werden. Eine Bedienungsanleitung des Almo-Programms zum Rasch-Modell ist enthalten in

"Wie man mit Almo ein Rasch-Modell rechnet.PDF"

(herunterladbar in der Handbuchseite von www.almo-statistik.de)

Das allgemeine ordinale Rasch-Modell wird in der Literatur auch als **partial credit model** oder abgekürzt **PCM** bezeichnet.

In **Almo** ist das ordinale Rasch-Modell als Prog14 mit den zwei Programm-Masken Prog14m3 und Prog14m4 enthalten. Es wird gestartet durch Klick auf den Knopf „Verfahren“ und dann „Rasch-Modell“. Das seither schon vorhandene *dichotome* Rasch-Modell mit den Programm-Masken Prog14m1 und Prog14ma ist nach wie vor gültig, ist jedoch im neuen ordinalen Rasch-Programm inkludiert. Die Ergebnis-Ausgabe ist weitgehend dieselbe. Das neue Programm ist jedoch vor allem bei den Modell- und Itemfit-Tests überlegen.

Der hier vorliegende Text beinhaltet folgendes:
- Es wird gezeigt, wie das Almo-Programm zum Rasch-Verfahren zu bedienen ist
- die Ergebnisse werden ausführlich anhand von simulierten und empirischen Beispielen erläutert
- die Theorie des Rasch-Modells wird überblicksweise dargestellt und seine Probleme diskutiert. Wer sich umfassender mit der Theorie (und auch den Problemen) des Rasch-Modells befassen

will, dem sei das Buch von Jürgen Rost (2004) empfohlen, das kompetent und verständlich das Modell darstellt.

**P14.8.1 Items für das ordinale Rasch-Modell**

Für die im ordinale Raschmodell zu skalierenden Variablen gelten folgende Möglichkeiten und Bedingungen:

1. In die Analyse dürfen dichotome und/oder polytome Items aufgenommen werden.
   Es können auch Analysen nur mit dichotomen oder nur mit polytomen Items durchgeführt werden.

2. Die Items dürfen ungleich viele Kategorien (Ausprägungen) besitzen

3. Das Rasch-Modell unterstellt, dass alle Items auf einer gemeinsamen Messdimension messen. Dies ist die "Annahme der Eindimensionalität"

4 Die Items besitzen unterschiedliche "Schwierigkeitswerte". Bei einem Wissenstest z.B. bedeutet dies, dass Wissensfragen unterschiedlich schwer zu "lösen" sind.

5. Die Items besitzen die gleiche Diskriminationsfähigkeit (Trennschärfe). Das heisst, sie trennen die Personen ungefähr gleich gut in jene, die "oberhalb" der Schwierigkeit des Items stehen und jene, die "unterhalb" stehen. Geometrisch bedeutet dies: Die Steigungen der logistischen Kurven aller Items müssen gleich sein.

6. Die Variablen müssen mit Einer-Schritten von ihrer Werte-Untergrenze zur Obergrenze kodiert sein. Die Werte-Untergrenze muss nicht notwendigerweise =0 sein. Ein Item mit 3 Kategorien kann z.B. mit `3,4,5` kodiert sein. Almo transformiert programmintern diese Werte automatisch auf Untergrenze 0. Ist die Bedingung der "Einer-Schritte" nicht erfüllt, dann müssen die Variablen in der Umkodierungs-Box entsprechend umkodiert werden

7. Besitzt ein Item 3 oder mehr Ausprägungen, dann müssen die Ausprägungen in ihrer Schwierigkeit ansteigend sein. Die den Item-Kategorien dann zugeordneten Zahlen sind in Einer-Schritten ansteigende ganzzahlige ordinale Ziffern

8. Die Variablen müssen in Richtung der Messdimension kodiert sein. Beispiel: Bei einem 0-1 kodierten dichotomen Item bedeutet die 1: liegt höher auf der Messdimension als 0. Ist dies abweichend bei einem Item nicht der Fall, dann muss dieses in seiner Kodierungsrichtung umgedreht werden

   Beispiel
   Die zu skalierende Variablen V1 bis V5 seien mit `3,4,5` kodiert,
   wobei allerdings bei Variable 5 die Kodierungs-Richtung umgekehrt ist als bei den anderen.
   Dann muss in der Umkodierungsbox geschrieben werden:

```
      V5( 5=3; 4=4; 3=5 )
oder V5( 3 bis 5 = Umdrehen)
```

**Beispiele für ordinal-polytome Rasch-Items**
Beispiel für Wissensfragen
In der Pisa-Studie werden den befragten Schülern viele Testfragen aus den Bereichen Mathematik, Wissenschaft, Lesen vorgelegt. Mit wenigen Ausnahmen werden diese Testfragen nicht veröffentlicht um sie bei nachfolgenden Befragungen wieder verwenden zu können. Bei Mathematik-Tests kann ein ordinal-polytomer Test dadurch entstehen, dass die Testleistung des Schülers beurteilt wird mit

(0) Aufgabe nicht gelöst
(1) teilweise gelöst
(2) ganz gelöst

Beispiel für Einstellungsfragen
Die Bereitschaft von Studenten zu nichtkonventionellen politischen Aktivitäten soll erfragt werden. Die Fragen lauten:

Würden Sie
Item 1: ... sich an einer Unterschriftenaktion beteiligen
Item 2: ... einen Brief an einen Politiker schreiben
Item 3: ... an einer Kundgebung teilnehmen
Item 4: ... in einer Bürgerinitiative mitarbeiten
Item 5: ... an der Besetzung eines öffentlichen Gebäudes teilnehmen

Als Antwort-Vorgaben für diese 5 Items wird den Studenten angeboten:

(0) nein, würde ich nicht tun
(1) vielleicht
(2) eher wahrscheinlich
(3) ja, würde ich tun

Es wird angenommen, dass die einzelnen Items unterschiedlich "schwierig" sind. Item 1 ist "leichter" als Item 2, da es eine geringere politische Aktivitätsbereitschaft erfordert. Und es wird erlaubt, dass der Befragte innerhalb eines Items nochmals eine Abstufung vornimmt.

**P14.8.2 Almo-Programm-Masken**
In Almo sind 2 Programm-Masken enthalten.

1. Prog14m3.Msk:
   Ordinales Rasch-Programm mit minimaler Ein- und Ausgabe

2. Prog14m4.Msk:
   Ordinales Rasch-Programm mit Optionen und Grafik

Daneben gibt es noch 2 Grafik-Programme, bei denen der Benutzer für frei gewählte Schwierigkeits- bzw. Schwellenwerte die Itemkurven von ordinal-polytomen Items erzeugen kann. Dies sind

1. Prog14g1.Msk
   Itemkurve im ordinalen Rasch-Model

2. Prog14g2.Msk
   Kategorienkurven im ordinalen Rasch-Model

2. Prog14g3.Msk
   Schwellenkurven im ordinalen Rasch-Model

Diese Programme werden durch Klick auf den Knopf "Verfahren" und "Rasch-Modell" erreicht. Weiterhin sind folgende Beispiel-Programme vorhanden

`Rasch_PisaMathe_A.Alm`. Das Prog rechnet mit den Mathe-Items aus Pisa 2009 für Austria
`Rasch_PisaMathe_A_D.Alm`. Das Prog rechnet mit den Mathe-Items aus Pisa 2009 für A+D
`Rasch_PisaMathe_EU.Alm`. Das Prog rechnet mit den Mathe-Items aus Pisa 2009
`Rasch_Neoffi.Alm`. Das Prog rechnet mit den „Neoffi-Daten“ (siehe Rost, 2004, S. 202)
`Rasch_KFT.Alm`. Das Prog rechnet mit den KFT-Daten (siehe Rost, 2004, S. 95)

Die Beispiel-Programme findet man nach Klick auf den Knopf "alle Progs" am Oberrand des Almo-Fensters. Sie können auch aus der Programm-Maske Prog14m4.Msk gestartet werden.

Die 3 Beispielprogramme aus Pisa 2009 wurden mit den Schülern gerechnet, denen die Mathe-Tests des "Booklet 1" zur Beantwortung tatsächlich vorgelegt wurden. D.h. es wurden keine Schüler einbezogen, deren Antworten (nach dem Pisa-Imputationsverfahren) "geschätzt" wurden.

### P14.8.3 Die logistische Itemfunktion im Rasch-Modell

Es wird folgende ***Notation*** verwendet:

Die Zeichen **`**`** bedeuten 'Potenzieren'
**`e`** = das ist die e-Zahl 2.718...
**`ap`** = Zahl der Ausprägungen der Variablen minus 1;
da die Nummerierung der Ausprägungen mit 0 beginnt
ist **`ap`** gleich der höchsten Ausprägungsnummer
**`j`** = das sind die Kategorien-Funktionen 0, 1, ... bis **ap**
**`i`** = Gesamtpunkt von 0 bis maxg (maximaler Gesamtpunkt)
**`maxg`** = maximaler Gesamtpunkt
**`f`** = Fähigkeitswert der Person (=Personenparameter)
**`f[i]`** = Fähigkeitswert, der dem Gesamtpunkt i entspricht
**`s`** = Schwierigkeitsparameter (=Itemparameter)
**`s[j]`** = Schwierigkeitsparameter für Kategorie j
**`w[j]`** = Schwellenwerte von Kategorie j
**`p`** = Wahrscheinlichkeit, dass aus den möglichen Kategorien des Items
die Kategorie j gewählt wird

Zur Terminologie: Die zentralen Begriffe des Rasch-Modells werden in unserer Darstellung mit

**Schwierigkeitsparameter**
und
**Fähigkeitsparameter**

bezeichnet. In der Literatur, z.B. bei Jürgen Rost (2004), werden häufig die neutraleren Begriffe

**Itemparameter**
und
**Personenparameter**

verwendet. Wir wollen jedoch bei unserer Terminologie bleiben, weil diese schon bei der Darstellung des dichotomen Rasch-Modells (in den vorausgegangenen Almo-Versionen) verwendet wurde und weil sie anschaulicher ist.
Der 3. zentrale Begriff des **Schwellenparameters** wird einheitlich verwendet. In der englischsprachigen Literatur wird von "threshold" gesprochen, was direkt übersetzt "Schwelle" heißt.

### P14.8.3.1 Die logistische Funktion im dichotomen Rasch-Modell

Das dichotome Rasch-Modell kann als eine Weiterentwicklung des Guttman-Verfahrens betrachtet werden. Aus der einstufigen "Treppenkurve" bei Guttman wird die S-förmige logistische Kurve bei Rasch. Im Almo-Handbuch "Sozialwissenschaftliche Skalierungsverfahren", Abschnitt P14.0 wurde das ausführlich dargestellt.

Die logistische Funktion des dichotomen Rasch-Modells für ein Item lautet

(0a) $$p = \frac{e^{(f-s)}}{1+e^{(f-s)}}$$

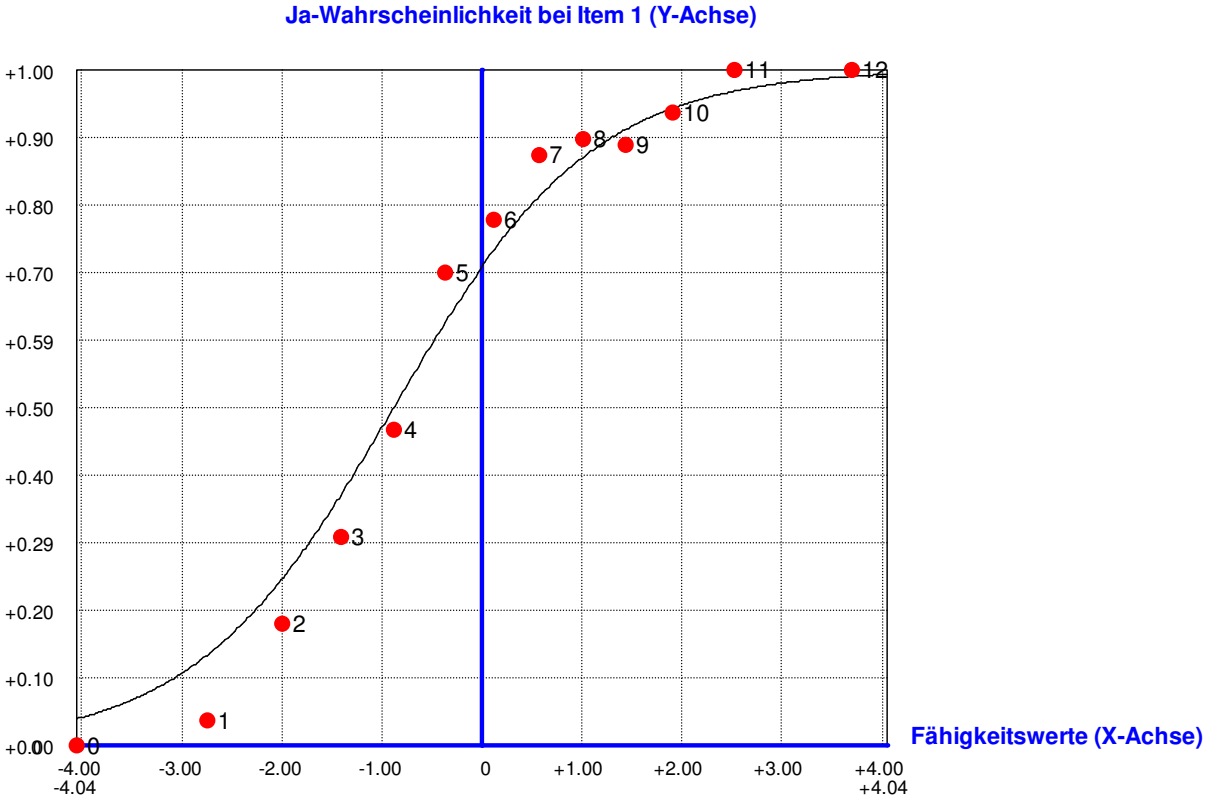
oder

(0b) $$p = \frac{1}{1+e^{(s-f)}}$$

```
p  ist die Wahrscheinlichkeit, dass die Person mit Gesamtpunkt i das
   Item positiv beantwortet

f  ist der Fähigkeitsparameter der Person mit Gesamtpunktwert i

S  ist der Schwierigkeitsparameter des Items.

Beachte: In Gleichung (0b) sind in der Klammer s und f gegenüber (0a)
vertauscht.
```

Die Funktion kann grafisch dargetellt werden. Wir wollen von folgender Datenlage ausgehen: Insgesamt wurden den Probanden 12 dichotome Items vorgelegt. Für eine positive Antwort erhält der Proband 1 Punkt. Maximal sind also 12 Gesamtpunkte erzielbar. Wir betrachten nun die

logistische Kurve von Item 1.




Die Formel für diese Kurve Nr. 1 ist `Y=1/(1+e**(-0.888-X))`. Der Parameter s ist -0.888. Die logistische Kurve im dichotomen Rasch-Modell ist eine S-förmige Kurve, die von links unten nach rechts oben verläuft.

Der rote Punkt 8 z.B. ist so zu interpretieren: Personen mit 8 Gesamtpunkten (aus den 12 möglichen) haben zu 90% (0.90 auf der y-Achse) das Item 1 positiv beantwortet; z.B. bei einem Mathe-Test die Aufgabe Nr. 1 gelöst. Ihre vom Rasch-Modell prognostizierte Wahrscheinlichkeit p ist mit 0.88 etwas kleiner. Ihr Fähigkeitswert f ist 1.00 (auf der x-Achse). Als Schwierigkeitswert des Items betrachtet man die Stelle auf der x-Achse, bei der die Wahrscheinlichkeit einer positiven Antwort 0.5 ist. An dieser Stelle liegt auch der Wendepunkt der Kurve. Das ist in obiger Grafik bei y=0.5 und x= -0.888. Die x-Achse ist die gemeinsame Skala der Fähigkeit f der Personen und der Schwierigkeit s der Items.

Auffällig an der logistischen Rasch-Funktion in Gleichung (0b) ist, dass sie keinen Parameter für die Steigung der Kurve besitzt. Die allgemeine logistische Funktion, die einen solchen Parameter besitzt – wir bezeichnen ihn mit c - lautet

(L1) $$y = \frac{1}{1+e^{-(c*(b+x))}}$$

Das negative Vorzeichen nach e kann in die Klammer hineingenommen werden. Die logistische Funktion lautet dann

(L2) $y = \frac{1}{1+e^{(c*(-b-x))}}$

Diese Formel wird im „Birnbaum“-Modell verwendet. Wir werden weiter unten darauf zurückkommen.

b bestimmt die horizontale Lage der Kurve. Je grösser b umso weiter links liegt die Kurve- bei positivem c. Ist c negativ dann umgekehrt. Bedeutsam ist die x-Position des Wendepunkts der Kurve

| | | |
|---|---|---|
| Der Wendepunkt liegt bei | (L2a) | `y = 0.5` |
| und | (L2b) | `x = b` |
| Die Steigung im Wendepunkt ist | (L2c) | `c` |
| Steigungswinkel in Grad: | (L2d) | `57.2957795*arctan(c)` |

Die horizontale Lage der logistischen Kurve ist durch den x-Wert des Wendepunkts bestimmt. Den Steigungswinkel im Wendepunkt kann der Benutzer mit dem in Almo enthaltenen „Kalkulator“ in folgender Weise berechnen: Mausklick auf das Menü „Statistik“ am Oberrand des Almo-Fensters. „Kalkulator“ selektieren. Obige Formel schreiben. Für c einen Wert eingeben. Klick auf den Knopf „Rechne“ in der Knopfleiste unterhalb der Menüleiste.

c bestimmt die Steilheit. Je grösser c absolut ist umso steiler. Vorzeichen positiv, dann wächst die Kurve von links nach rechts, Vorzeichen negativ, dann umgekehrt

Beim Skalierungsverfahren nach Rasch wird

1. x als Fähigkeitsparameter der Person und b (vorläufig) als "Leichtigkeits"-Parameter des Items interpretiert.
2. Der die Obergrenze bestimmende Parameter a wird auf 1 gesetzt, da eine Wahrscheinlichkeit nicht grösser als 1.0 werden kann
3. Der die Steilheit bestimmende Parameter c wird inhaltlich als die Trennschärfe des Items interpretiert. c wird im Rasch-Modell auf 1 gesetzt - könnte aber ohne Auswirkung auch auf einen anderen Wert, z.B. 4 oder 0.5 gesetzt werden. Entscheidend ist jedoch, dass c für alle Items einer Item-Batterie den gleichen Wert haben muss. Die Trennschärfe der Items muss gleich sein. Das ist eine für das Rasch-Modell sehr zentrale Bedingung. Durch die Maße "Infit" und "Outfit" wird diese Bedingung getestet. Siehe Abschnitt P14.8.8.4.1.

Werden der Parameter c auf 1 gesetzt, dann wird aus Gleichung (L2)

(L3) $y = \frac{1}{1+e^{(-b-x)}}$

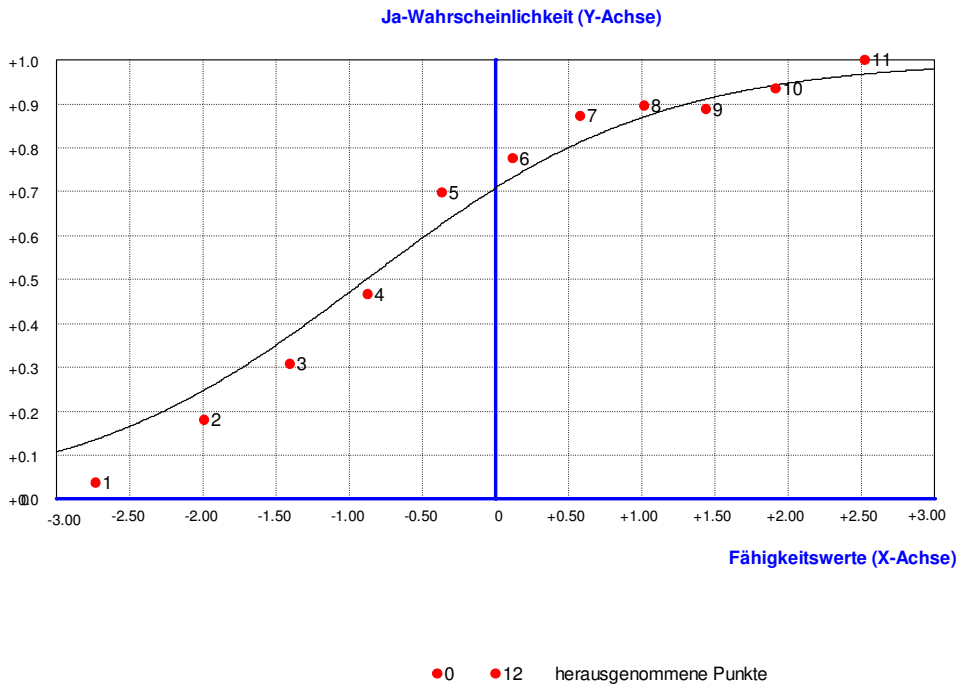
Im Rasch-Modell interessiert jedoch nicht die "Leichtigkeit" des Items, sondern das Gegenteil, seine Schwierigkeit. So wird in der Gleichung üblicherweise das negative Vorzeichen von b zum positiven. Die Gleichung lautet dann

(L4) $y = \frac{1}{1+e^{(b-x)}}$

Das ist obige Gleichung (0b). Wir haben oben b mit s und x mit f bezeichnet. In Gleichung (0b) bzw. (L4) ist b (bzw. s) so zu interpretieren:

Je grösser die Schwierigkeit b des Items umso weiter rechts liegt die Kurve. Die Kurve beginnt bei y=0 und erreicht mit y=1 ihren höchsten Wert. Die beiden extremen y-Werte werden erst bei x gleich minus unendlich bzw. plus unendlich erreicht. Die Kurve steigt von links nach rechts. Der Wendepunkt ist (gemäß Gleichung L2a) in diesem Falle bei y=1/2 und (gemäß L2b) x=b. Die Steigung im Wendepunkt ist (gemäß L2c) 1.0. Die Tangente am Wendepunkt y=0.5 steigt mit 1.0; das sind (gemäß L2d) 45 Grad.

Obige Grafik ist also verzerrt gezeichnet. Almo hat die y-Achse zu lang gezeichnet, wodurch der Winkel steiler wird. Das ist absichtlich geschehen, um den S-förmigen Verlauf der logistischen Funktion deutlich sichtbar zu machen. Wird die Grafik mit der korrekten Steigung von 45 Grad gezeichnet, dann entsteht:




Die minimale Punkt 0 und der maximale Punkt 12 wurden herausgenommen. Sie sollten nicht interpretiert werden. Der S-förmige Verlauf der logistischen Kurve ist nicht so gut zu erkennen wie bei der obigen verzerrten Abbildung.

Wenn der Benutzer die Grafik in verzerrter Form (wie oben) zeichnen will, dann muss er in der Programm-Maske Prog14m4 die Optionsbox „Grafik-Optionen“ öffnen und im 3. Eingabefeld eine 1 einsetzen. Siehe Abschnitt P14.8.7.12.

Dadurch dass das Rasch-Modell c auf 1 setzt, wird die logistische Funktion drastisch vereinfacht und es wird möglich die Parameter b und x inhaltlich zu interpretieren als Schwierigkeits- und Fähigkeitsparameter. Selbstverständlich sind Modelle entwickelt worden, die diese Vereinfachung

nicht vornehmen. Besonders bekannt wurde das Birnbaum- Modell, das die Steigung c nicht festlegt. Das verursacht jedoch fast unlösbare Interpretationsprobleme. Das Modell lässt dann zu, dass sich die Kurven verschiedener Items schneiden, d.h. dass besonders fähige Personen, die schwere Items lösen, systematisch bei leichten Items versagen. Die grundlegenden Konzepte des Schwierigkeits- und des Fähigkeits-Parameters werden damit unklar. Zur Diskussion des Birnbaum-Modells siehe z.B. Rost (2004, S. 133 und 370).
Siehe auch Abschnitt P14.8.8.4.2.1 „Die Steigung der logistischen Rasch-Funktion“.

### P14.8.3.2 Die Kategorienfunktion im allgemeinen ordinalen Rasch-Modell

Die Formel für die Kategorien-Kurven beim polytom-ordinalen Raschmodell läßt sich nicht so einfach und anschaulich darstellen wie die logistische Funktion für die Kurven-Formel beim dichotomen Rasch-Modell.

Wenn ein Item beispielsweise 4 Ausprägungen besitzt, dann gibt es die 4 Kategorien-Kurven j=0, =1, =2, =3

Die Formel für die Kategorien-Funktion lautet :

**(1)** $$p = \frac{e^{(j*f[i]-s[j])}}{sum}$$

dabei ist **sum**

**(2)** $$sum = \sum_{k=0}^{ap} e^{(k*f[i]-s[k])}$$

```
i     ist der Gesamtpunkt
j     ist die Kategorie (die Ausprägung) des Items (beginnend
      mit j=0)

ap    Zahl der angebotenen Kategorien des Items minus 1
k     Index, der über die Zahl der Ausprägungen des Items von k=0 bis
      k=ap läuft. Die Summierung in (2)läuft also über die Zahl der
      Ausprägungen des Items von k=0 bis k=ap

f[i]  Fähigkeitswert der Person mit i Gesamtpunkt
s[j]  Schwierigkeitswert der Kategorie j des Items

p     ist die Wahrscheinlichkeit, dass die Person mit Gesamtpunkt i die
      Kategorie j aus den vorhandenen ap+1 Kategorien des Item auswählt

      oder "anders herum" formuliert:
      ist die Wahrscheinlichkeit, dass aus den vorhandenen ap+1
      Kategorien die Kategorie j des Items von der Person mit i
      Gesamtpunkten gewählt wird.
```

Empirisch gegeben sind j und i. Vorgegeben ist die logistische Funktion gemäß (1) und (2). Die Aufgabe besteht nun darin, die f und s zu schätzen.

Diese Formeln gelten auch für dichotome Items mit den Ausprägungen 0 und 1.

**sum** wird dabei $sum = e^{(0*f[i]-s[0])} + e^{(1*f[i]-s[1])}$
bzw.

$$sum = 1 + e^{(1*f[i]-s[1])}$$

und

$$p\ (f\text{ür}\ Kategorie\ j = 0) = \frac{1}{1+e^{(1*f[i]-s[1])}}$$

$$p\ (f\text{ür}\ Kategorie\ j = 1) = \frac{e^{(1*f[i]-s[1])}}{1+e^{(1*f[i]-s[1])}}$$

bzw.

$$p\ (f\text{ür}\ Kategorie\ j = 1) = \frac{1}{1+e^{(s[1]-f[i])}}$$

Vergleiche oben mit Gleichung (0a)

In Gleichung (1) wird ausgedrückt,

1. dass eine Person mit einem Fähigkeitswert f
   (der dem Gesamtpunktwert i aus der Beantwortung aller Items entspricht)
2. aus den angebotenen ap Kategorien die Kategorie j, die einen
   Schwierigkeitswert s besitzt mit der Wahrscheinlichkeit p wählt

Nachfolgendes C-Programm ermittelt die p-Werte aller maxg Fähigkeitswerte für alle Kategorien-Kurven 0 bis ap

Liegen anstelle der Schwierigkeitsparameter s die Schwellenparameter w vor, so müssen die Schwellenparameter w zuerst in die Schwierigkeitsparameter s umgeformt werden – in folgender Weise:

```
s[0] = w[0] = 0;
for(j=0; j<=ap; j++)  s[j+1] = s[j]+w[j];
for(j=1; j<=ap; j++)  s[j]=s[j+1];
```

Die p-Werte der Kategorienfunktion über dem Wert f[i] ergeben sich dann so

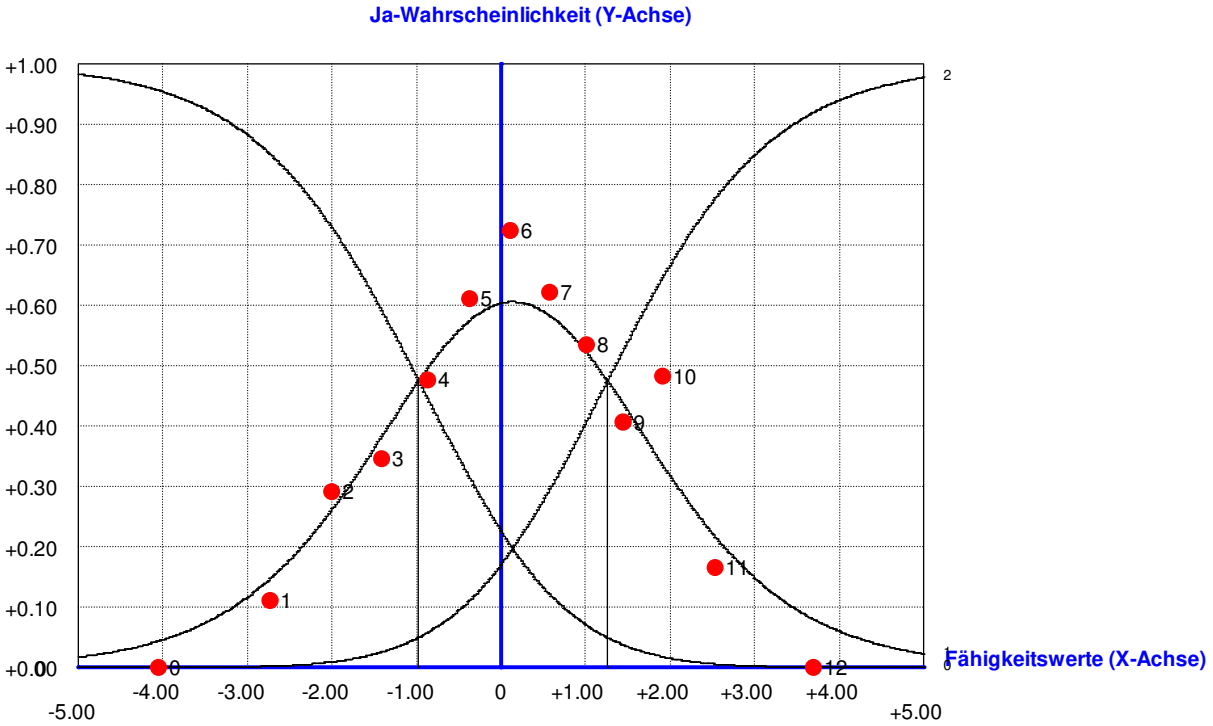
```
s[0]=0;
for(j=0; j<=ap; j++) begin        // laeuft ueber die Auspraegungen 0 bis ap
   for(i=0; i<=maxg; i++) begin // laeuft ueber die Gesamtpunkte 0 bis maxg

      sum=.0;
      for(k=0; k<=ap; k++) sum = sum + exp( k*f[i]-s[k] );
```

```
        p = exp( j*f[i] - s[j] ) / sum ;
    endfor
endfor
```

Zur Kurvenformel siehe J. Rost: Lehrbuch: Testtheorie - Testkonstruktion, Huber 2004, S.209, Gleichung 8 und 11

**Die Kategorienkurven eines polytomen Items**

Betrachten wir einen Item mit den 3 Kategorien 0, 1 und 2



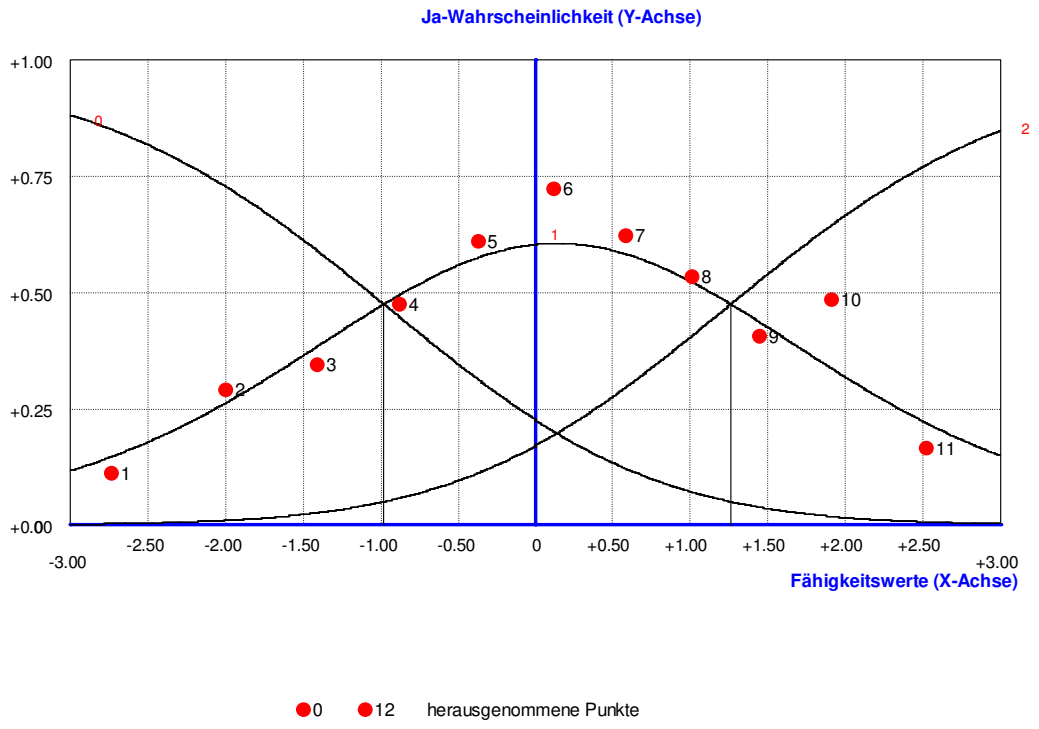

Die Grafik enthält entsprechend den 3 Katagorien 3 Kurven. Kurve 0 verläuft von links oben nach rechts unten. Kurve 1 ähnelt einer „Glockenkurve“. Kurve 2 ist eine logistische Kurve wie bei den dichotomen Items. Kurve 0 ist eine "umgedrehte" logistische Kurve. Eine Besonderheit ist die Kurve 1. Sie besitzt ein Maximum (kleiner 1) und läuft dann wieder nach unten.

Die 3 Kategorienfunktionen weisen folgende Merkmale auf. Die eingezeichneten Punkte werden anschliessend diskutiert.

1. Für Personen mit Fähigkeitswerten von -unendlich bis -0.98 ist die Wahrscheinlichkeit, dass sie von den 3 alternativen Kategorien des Items die Kategorie 0 wählen am grössten. Von Fähigkeit -0.98 bis +1.26 ist die Wahrscheinlichkeit für die Ausprägung 1 am grössten und von +1.26 bis plus unendlich die für Ausprägung 2. Die 2 Punkte -0.98 und +1.26 werden „Schwellenparameter“ oder "Schwellenpunkte" genannt. Das Kontinuum von minus bis plus unendlich wird in 3 Bereiche untergliedert. Jede Kategorie besitzt ihren Bereich, in dem sie die dominierende Wahrscheinlichkeit besitzt. Die x-Achse ist also die gemeinsame Skala, auf der die Fähigkeit der

Personen und die Schwellenwerte der Kategorien des Items abgetragen werden.

2. Über jedem Punkt der x-Achse addieren sich die p-Werte, d.h. die Wahrscheinlichkeiten der 3 Kurven zu 1.0.
Wie die Kurve für das dichotome Item, so ist auch obige Grafik stark verzerrt. Die Kurven 0 und 2 sind logistische Kurven (wie bei den dichotomen Items). Die Steigung im Wendepunkt der Kurve 2 ist (gemäß Gleichung L2c) 1.0 (=45 Grad), die der Kurve 0 ist -1 (=-45 Grad). Wird die Grafik entzerrt, dann entsteht folgendes:




Die Punkte 0 und 12 wurden herausgenommen um die x-Achse zu verkürzen.
Wenn der Benutzer die Grafik in entzerrter Form (wie hier) zeichnen will, dann muss er in der Programm-Maske Prog14m4 die Optionsbox „Grafik-Optionen“ öffnen und im 2. Eingabefeld eine 0 einsetzen. Siehe Abschnitt P14.8.7.12. Siehe auch Abschnitt P14.8.8.4.2.1 Die Steigung der logistischen Rasch-Funktion.

**Die in die Grafik eingezeichneten Punkte.**
Dies sind Gesamtpunktwerte. In obiger Grafik wurden jene Personen herausgegriffen, die die Kategorie 1 gewählt haben. Betrachten wir Punkt 3. Insgesamt haben 84 Personen 3 Gesamtpunkte erzielt. Davon haben sich beim betrachteten Item 29 für Kategorie 1 entschieden. Das ist ein Anteil von 29/84=0.35. Das ist ihre tatsächliche empirische Wahrscheinlichkeit, die Ausprägung 1 zu wählen. Der entsprechende Punkt auf der Kurve ist die vom Rasch-Modell prognostizierte "theoretische Wahrscheinlichkeit" von 0.38. Die Gesamtpunkte liegen sehr dicht um die Kurve herum - vielleicht mit Ausnahme von Punkt 6.

In die Grafik der 3 Kurven werden in Almo dann noch die Gesamtpunkte der Personen

eingezeichnet, die sich für Ausprägung 0 und 2 entschieden haben. So entstehen für das Item insgesamt 3 Grafiken. Das könnte auch in einer einzigen Grafik gemacht werden (mit verschiedenen Farben) - könnte aber unübersichtlich werden. In den drei Kurven liegen die Punkte alle sehr dicht bei den Kurven, so dass gefolgert werden kann: Das Item ist "Rasch-konform". Das ist natürlich ein intuitiver, dafür aber sehr anschaulicher Itemfit-Test. Mit den Koeffizienten "Outfit" und "Infit" und weiteren Koeffizienten werden in Almo standardisierte Tests angeboten, die allerdings ihre Probleme haben. Siehe Abschnitt P14.8.7.6.

Anmerkung: Wenn Personen dieses Item mit 1 beantwortet haben, dann kann ein Gesamtpunkte von 0 nicht auftreten, da sie ja in diesem Item bereits einen Punkt erzielt haben. Es ist auch nicht möglich, dass sie den maximalen Gesamtpunkt von 12 erreichen, da sie ja in diesem Item nicht den Höchstwert 2 erzielt haben. Diese beiden Punkte werden aus der Grafik herausgenommen und von Almo unterhalb der Grafik platziert, so dass der Benutzer auf einen Blick erkennt, welche Punkte "unlogisch" sind. In obiger Grafik wurde das nicht gemacht.

### P14.8.4 Schätzung der Parameter

Für die oben angegebenen Gleichungen (1) und (2) besteht nun die Aufgabe darin, die Schwierigkeitsparameter **s** (bzw. die Schwellenparameter **w**) und die Fähigkeitsparameter **f** zu schätzen.
In Almo werden die "Schwierigkeitsparameter" geschätzt. Durch eine einfache Transformation gehen aus Ihnen die "Schwellenwerte" hervor

Für **s** bzw. **w** existieren 3 Schätzverfahren:

1. die unbedingte Maximum-Likelihood-Schätzung (kurz: **uML** oder auch **JML**)
2. die bedingte (conditioned) Maximum-Likelihood-Schätzung (kurz: **cML**)
2. die marginale Maximum-Likelihood-Schätzung (kurz: **mML**)

Im Vergleich von **uML** und **cML** gilt letztere allgemein als das überlegene Verfahren. Almo verwendet dieses. Die beiden Verfahren liefern unterschiedliche Schwierigkeits- bzw. Schwellenparameter - wobei die Unterschiede zwar deutlich aber nicht sehr groß sind. Neben Almo verwenden auch die Programme **Winmira** (von Davier) und **Multira** (Carstensen und Rost) die **cML**-Schätzung. Bei gleicher Kategorienzahl der Items liefern diese Programme exakt identische Ergebnisse. Bei ungleicher Kategorienzahl können zwischen diesen 3 Programme unterschiedliche Ergebnisse auftreten, die allerdings bedeutungslos sind. Sie bestehen immer nur in einer additiven Konstanten - in gleicher Höhe bei den Schwierigkeits/Schwellen- und den Fähigkeitsparametern. Siehe dazu den nachfolgenden Abschnitt.

Die **uML** wird vom Programm **Winsteps** verwendet (dort **JML** =joint ML genannt, da bei diesem Verfahren gleichzeitig Item- und Personenparameter ermittelt werden).

Die marginale **mML**-Schätzung unterstellt, dass die Fähigkeitswerte der Personen einer bestimmten Verteilung (in der Regel der Normalverteilung) folgen (Rost, 2004, S. 309, ConQuest-Handbuch, S. 135). Dieses Schätzverfahren wird im Programm **ConQuest** verwendet, wobei die Programm-Autoren eigene Adaptionen und eigenständige Entwicklungen habe einfliessen lassen. Siehe dazu das ConQuest-Handbuch, Kapitel 12 von Wu, Adams, Wilson, Haldane. Das Programm ist (für ein wissenschaftliches Programm) enorm teuer. Das Handbuch kann allerdings im Internet frei heruntergeladen werden.

Für die Schätzung der Fähigkeitsparameter existieren 2 Verfahren

1. die unbedingte Maximum-Likelihood-Schätzung (kurz: **uML** oder **ML** oder **JML**)
2. die gewichtete Maximum-Likelihood-Schätzung nach Warm (kurz: **WLE**)

Übereinstimmend wird die **WLE** ("weigthed likelihood estimate") als das überlegene Verfahren betrachtet. In den 4 Programmen Almo, Conquest, Multira, Winmira wird (teilweise zusätzlich zu ML) die WLE-Schätzung verwendet.

Die Schätzmethoden werden übersichtlich und vergleichend dargestellt bei Jürgen Rost, 2004, Kap. 4.2 und S. 127.

Die in Almo verwendeten Schätzverfahren werden im Anhang von Heinrich Potuschak in aller Ausführlichkeit dargestellt. Die entsprechenden Unterprogramme in C und R sind im Almo-Ordner "**Algorithmen_in_C**" und „**Rasch_in_R**“ enthalten. Sie können frei benutzt werden.

**P14.8.5 Vergleich von Almo, Conquest, Multira, Winmira**
Es wurden die ersten 5 Items aus den **NEOFFI**-Daten von Borkenau/Ostendorf (1991) gerechnet. Die 5 Items besitzen 4 Kategorien. Siehe Rost, 2004, S.202 ff

Itemparameter (Schwellenparameter)

| | | Almo Winmira | Multira | ConQuest |
|---|---|---|---|---|
| V1 | 1 | -3.662 | -3.657 | -3.660 |
| | 2 | 0.144 | 0.142 | 0.147 |
| | 3 | 1.779 | 1.777 | 1.803 |
| V2 | 1 | -2.010 | -2.013 | -2.069 |
| | 2 | 0.477 | 0.474 | 0.499 |
| | 3 | 2.114 | 2.111 | 2.129 |
| V3 | 1 | -2.410 | -2.413 | -2.454 |
| | 2 | 1.203 | 1.200 | 1.225 |
| | 3 | 2.508 | 2.505 | 2.507 |
| V4 | 1 | -3.879 | -3.849 | -3.867 |
| | 2 | 0.065 | 0.065 | 0.067 |
| | 3 | 2.150 | 2.148 | 2.169 |
| V5 | 1 | -2.087 | -2.089 | -2.135 |
| | 2 | 1.340 | 1.337 | 1.365 |
| | 3 | 2.267 | 2.264 | 2.275 |

Die Itemparameter in Almo, Winmira, Multira entstehen aus cML. Winmira und Almo sind völlig identisch. Die Werte von Multira zeigen erst an der 3. Kommastelle andere Zahlen. Solche

Unterschiede treten wohl auf, wenn Programme nach dem gleichen Kalkül in verschiedenen Programmiersprachen geschrieben sind und vorgefertigte aber unterschiedliche mathematische Standard-Prozeduren mit unterschiedlichen Genauigkeiten verwenden.

Die Schwellenparameter in ConQuest werden als "item deltas" bezeichnet, die im Rahmen der "traditional Item analysis" gewonnen werden. Nach ihnen werden auch die Kurvenplots gezeichnet. Die Parameter erhält man durch Klick auf "Tables/Write Delta File". Die Kurvenplots erhält man durch Klick auf "Plot/Characteristic Curves by Category" Die "item deltas" in ConQuest unterscheiden sich erst an der 2. Kommastelle von denen der anderen Programme, was natürlich auf das andere Schätzverfahren zurückzuführen ist.

Die Itemparameter, die man in ConQuest durch Klick auf "Tables/Item Parameter Estimates" erhält, entstanden aus der mML-Schätzung. Sie sind für jedes einzelne Item auf Summe 0 normiert, was durch den speziellen Kalkül von Wu, Adams, Wilson, Haldane erzwungen wird. Sie sind mit den gängigen Schwellenwerten überhaupt nicht vergleichbar. Siehe dazu das ConQuest-Handbuch, S.33. Sie sind jedenfalls nicht substantiell als Schwierigkeits– oder Schwellenwerte interpretierbar.

Personenparameter (Fähigkeitswerte)

| Ges.pkt | Almo Winmira | Multira | ConQuest |
|---|---|---|---|
| 0 | -5.553 | -5.538 | -5.557 |
| 1 | -4.127 | -4.117 | -4.140 |
| 2 | -3.231 | -3.225 | -3.254 |
| 3 | -2.466 | -2.464 | -2.497 |
| 4 | -1.740 | -1.740 | -1.769 |
| 5 | -1.016 | -1.017 | -1.031 |
| 6 | -0.359 | -0.361 | -0.359 |
| 7 | 0.186 | 0.184 | 0.195 |
| 8 | 0.649 | 0.647 | 0.662 |
| 9 | 1.063 | 1.060 | 1.077 |
| 10 | 1.451 | 1.449 | 1.467 |
| 11 | 1.837 | 1.834 | 1.852 |
| 12 | 2.244 | 2.241 | 2.259 |
| 13 | 2.715 | 2.712 | 2.728 |
| 14 | 3.342 | 3.340 | 3.354 |
| 15 | 4.544 | 4.541 | 4.555 |

Die Personenparameter aller Programme enstehen aus WLE

Das für den empirischen Forscher wichtigste Ergebnis sind sicherlich die Personenparameter. Diese will er seinen Untersuchungsobjekten zuweisen um damit dann weitere statistische Analysen zu rechnen. Hier erweist es sich nun, dass die 4 Programme praktisch dasselbe Ergebnis liefern.

Almo und Winmira sind wieder vollkommen identisch. Die Unterschiede zu ConQuest und Multira treten erst an der 2. Kommastelle auf und sind somit vernachlässigbar.

Zwischen Almo und Winmira besteht bei einer besonderen Konstellation doch ein Unterschied: Bei ungleicher Kategorienzahl der Items können unterschiedliche Ergebnisse auftreten. Diese bestehen immer nur in einer (geringen) additiven Konstanten - in gleicher Höhe bei den Item- und den Personenparametern und sind dadurch bedeutungslos. Wird zu den Itemparametern und den Personenparametern eine (für beide gleiche) Konstante addiert, so hat das keine Auswirkung auf die vom Modell geschätzte Lösungs-Wahrscheinlichkeit **p**. Die Unterschiede kommen vermutlich durch unterschiedliche Normierungen der Itemparameter zustande. In Almo werden die Itemparameter in allen Fällen strikt auf Gesamtsumme 0 normiert; in Winmira und Multira in diesem Falle nicht.

### P14.8.6 Eingabe in die Programm-Maske Prog14m3

Prog14m3 wird geladen durch Klick auf den Knopf "Verfahren". Dann "Rasch-Modell" oder "Skalierungsverfahren". Nach Klick auf den Knopf "Rechne" wird gerechnet. Siehe auch das PDF-Dokument „BedienungsAnleitung_RaschModell.PDF“ in

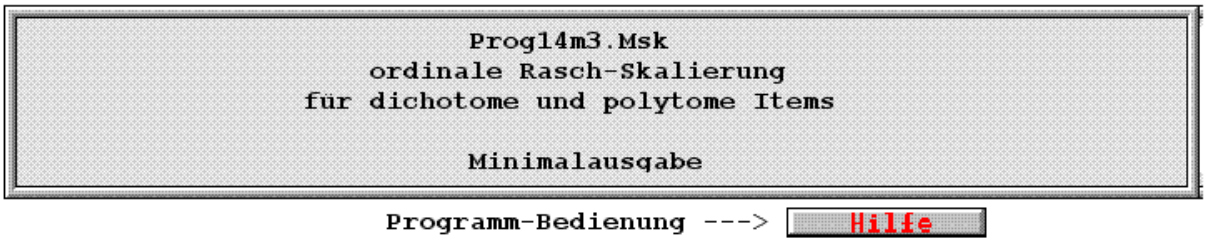
www.almo-statistik.de/handbuch.html

Fast alle nachfolgenden Eingabeboxen sind in aller Ausführlichkeit im Handbuch

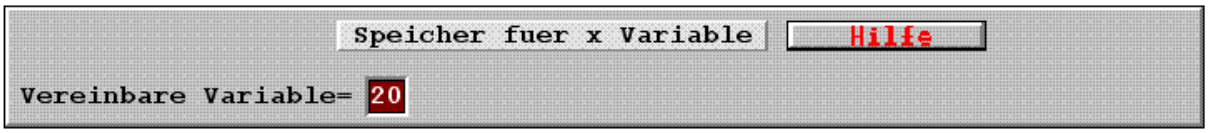
**"P0 Arbeiten mit Progs"**

erklärt - so dass sie hier relativ kurz erläutert werden können.

**P14.8.6.1** Programm-Kopf




**P14.8.6.2** Speicher

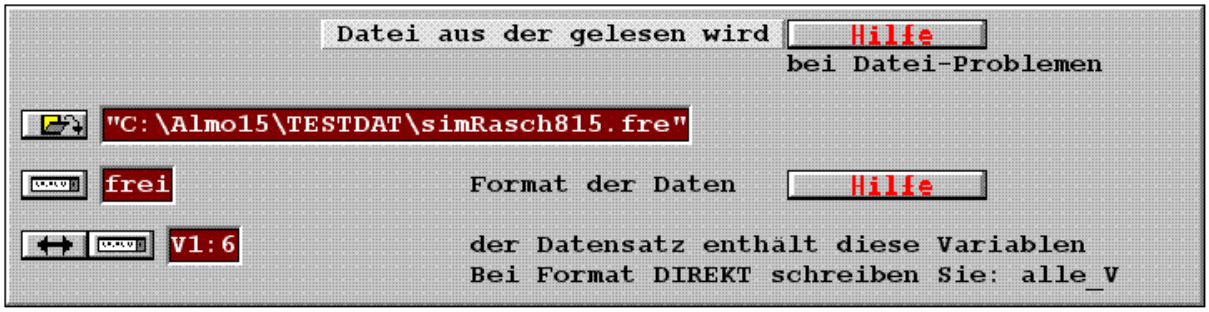


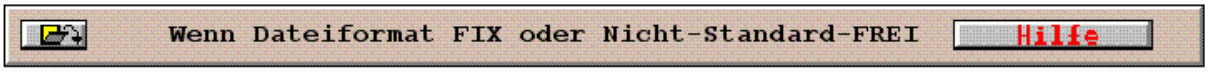

Geben Sie mindestens soviele Variable an, wie der Datensatz umfasst; eher mehr.
Siehe dazu das oben erwähnte Handbuch, Abschnitt P0.1

**P14.8.6.3** Vereinbarungen

Diese Eingabebox nur öffnen, wenn Almo nach einem 1. Programmlauf dazu auffordert.
Siehe dazu das oben erwähnte Handbuch, Abschnitt P0.2

**P14.8.6.4** Daten einlesen




Geben Sie hier den Namen der Datei an (den vollen Pfad- und Dateinamen), in dem sich die zu analysierenden Daten befinden. Die häufigsten Probleme, die Benutzer mit Almo haben treten an dieser Stelle auf. Lesen Sie insbesondere die Hilfe zum "Format". Siehe dazu das oben erwähnte Handbuch, Abschnitt P0.3

In Prog14m3 werden als Beispiel die Daten "simRasch815.fre" nach dem Rasch-Modell ausgewertet. Die Daten wurden nicht empirisch gewonnen. Sie wurden simuliert. Es wurden 815 Datensätze erzeugt. Die ersten und die letzten Datensätze sind folgende

```
0 0 1 1 1 0
1 0 0 0 1 1
0 0 0 1 1 0
0 0 0 0 1 1
1 0 0 1 1 0
1 0 0 0 1 0
. . . . . .
. . . . . .
. . . . . .
1 0 2 2 3 1
1 1 1 2 2 3
1 0 2 2 1 3
1 1 2 1 2 3
1 1 1 2 2 3
1 1 1 1 3 3
```

Die Daten befinden sich im Format FREI. Die Zahlenwerte sind durch ein Blank (oder Komma) getrennt. V1 und V2 sind dichotom. V3 und V4 besitzen 3 Ausprägungen. V5 und V6 besitzen 4 Ausprägungen. Alle Variable sind mit 0 beginnend in Einer-Schritten kodiert. Sie können die Daten anschauen, wenn Sie in der Dateibox auf den Öffne-Knopf vor dem Dateinamen klicken oder in der Knopfleiste am Oberrand des Almo-Fensters den Knopf "Zeige Daten" klicken.

Empirische Daten enthalten in der Regel "fehlende Werte" (in Almo auch "KeinWert" genannt). Ist für einen Item "KeinWert" mit einem vom Benutzer festgelegten Wert kodiert (z.B. mit der Zahl

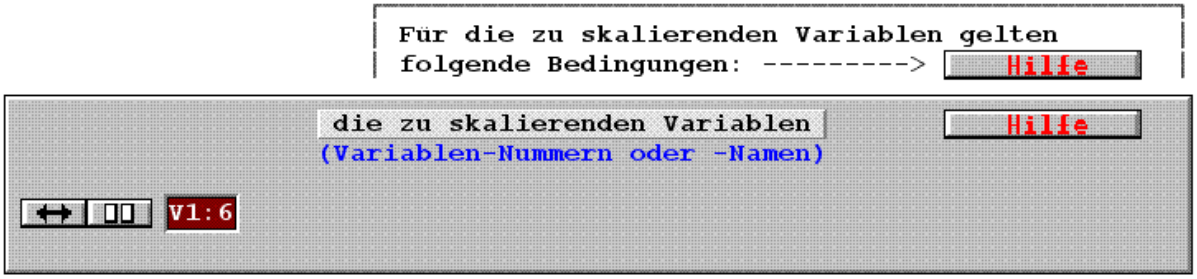
99) dann muss in der nachfolgend beschriebenen "Umkodierungsbox" eine entsprechende Deklaration erfolgen, z.B. so

```
V4 (99 = KeinWert)
```

Siehe in der Umkodierungsbox den Hilfeknopf "wie man die Kein-Wert-Angabe schreibt"

Wenn die Daten in einem Fremdformat vorliegen, dann ist es möglich sie in ein Almo-Format zu transformieren. Der Benutzer sollte dazu lesen: Almo-Handbuch Teil 1 Bedienungsanleitung, Kapitel 8.

**P14.8.6.5** Items auswählen

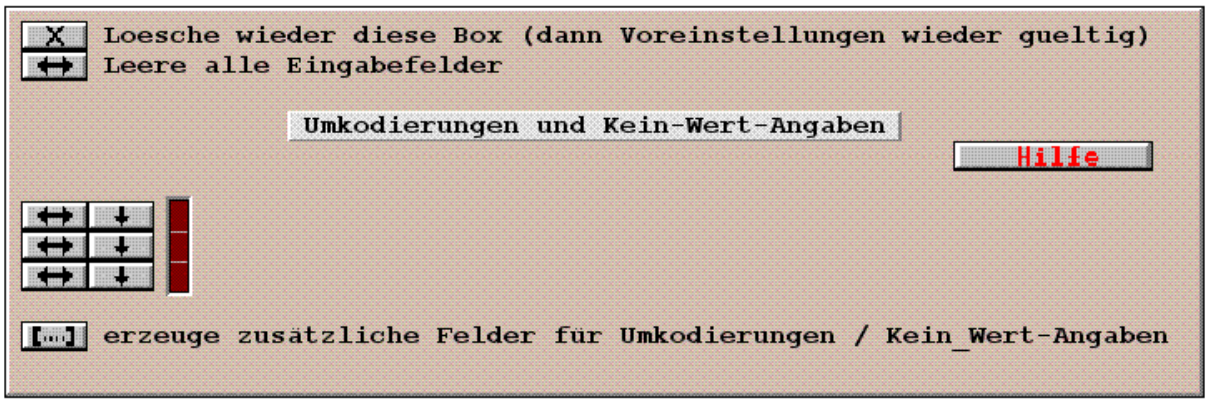



Geben Sie die Nummern der Variablen ein, die dem Rasch-Verfahren unterworfen werden sollen. Die Reihenfolge ist beliebig. Die Bedingungen, die die Variablen erfüllen müssen, können durch Klick auf den Hilfeknopf oder oben in Abschnitt P14.8.1 nachgelesen werden.

**P14.8.6.6** Umkodierungen

Wenn die Variablen nicht den Bedingungen entsprechen, die für eine Rasch-Analyse erfüllt sein müssen, dann können diese in der Regel durch entsprechende Umkodierungen hergestellt werden.

Die geöffnete Umkodierungsbox ist folgende




Klicken Sie auf den Hilfeknopf und lesen Sie wie in Almo umkodiert wird.

**Ausgabe aus Prog14m3**

Die Ausgabe besteht aus mehreren Teilen, die auch im nachfolgend beschriebenen Ergebnis von Prog14m4 enthalten sind. Sie werden dort im Detail erklärt.

### P14.8.7 Eingabe in die Programm-Maske Prog14m4

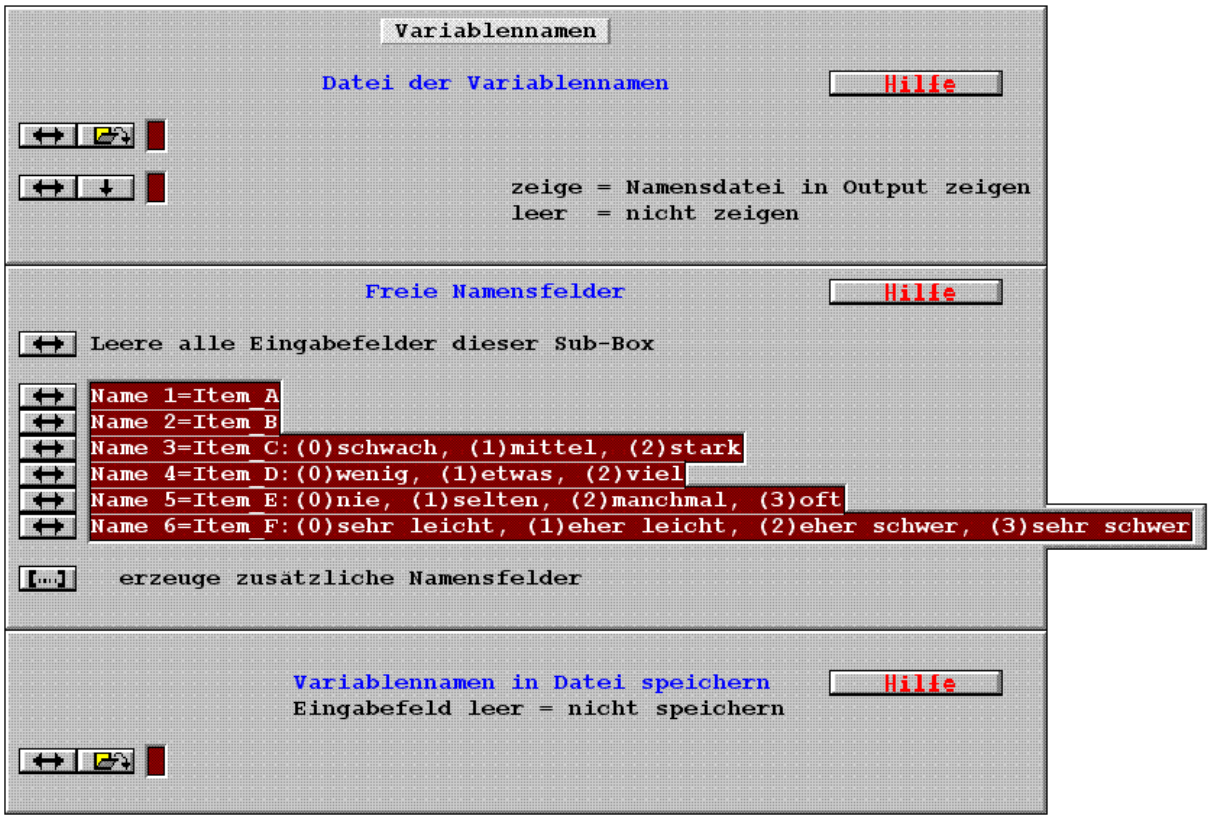
Siehe auch das PDF-Dokument „BedienungsAnleitung_RaschModell.PDF“ in

www.almo-statistik.de/handbuch.html

Fast alle nachfolgenden Eingabeboxen sind in aller Ausführlichkeit im Handbuch "**P0 Arbeiten mit Progs**" ausführlich beschrieben.

Die beiden Eingabeboxen "*Speicher für x Variable*" und "*Option: Weitere Vereinbarungen ....*" wurden bereits oben abgebildet.

#### P14.8.7.1 Variablennamen




Es müssen keine Variablennamen eingegeben werden. Almo verwendet in diesem Fall automatisch und selbsttätig die Namen 'V1', 'V2' usw. Diese werden vom Benutzer also nicht geschrieben.

Diese Eingabebox für Variablennamen bietet mehr Möglichkeiten als die oben bei Prog14m3 abgebildete. Der Benutzer kann Variablennamen in die 2. Subbox einschreiben oder eine Datei

aufrufen, in die er zuvor Variablennamen geschrieben hat. Lesen Sie die Hilfen, die mit Klick auf den jeweiligen Hilfeknopf angezeigt werden und die ausführlichen Hilfen im oben angegebenen Handbuch, Abschnitt P0.3.

Im verwendeten Beispiel wurden Variablen- und Ausprägungsnamen in die 2. Subbox geschrieben. Die ersten beiden Variablen wurden "Item_A" und "Item_B" genannt. Ihre Ausprägungen haben keine Namen bekommen. Die nächsten beiden Variablen besitzen 3 Ausprägungen, die 0-1-2 kodiert sind. Für Item_C wurden die Ausprägungsnamen (0)schwach, (1)mittel, (2)stark verwendet usw. BEACHTE: Wären die Ausprägungen z.B. mit 2-3-4 kodiert, dann müssten diese Nummern in die Klammern vor den Ausprägungsnamen geschrieben werden. Oder sehr viel einfacher: Man gibt die Ausprägungsnamen ohne vorausgehende Nummern und Klammern an. Almo ordnet dann die Ausprägungsnamen nacheinander den wachsenden Zahlenwerten zu egal mit welcher Zahl diese beginnen. Siehe dazu P0.3, Abschitt "Wie man Ausprägungsnamen schreibt".

Die Eingabebox für die Daten und die ergänzende Box für den Fall des Dateiformats FIX oder Nicht-Standard-Frei wurde oben bereits dargestellt

Auch die Box für die zu skalierenden Variablen (auf die das Rasch-Verfahren) angewendet werden soll) wurde oben gezeigt. Wurden den Variablen Namen gegeben, dann können hier wahlweise Variablen-Nummern oder Namen (nicht Ausprägungsnamen) eingegeben werden.

**P14.8.7.2** Ein- und Ausschliessen von Untersuchungseinheiten

Wird die Box geöffnet, dann sieht man folgendes:

Loesche wieder diese Box (dann Voreinstellungen wieder gueltig)
Leere alle Eingabefelder

Ein- und Ausschliessen von Untersuchungseinheiten

Hilfe
Hilfe

nur diese Datensätzen verwenden
wenn Eingabefeld leer, dann alle
Hilfe

Untersuchungseinheiten in Analyse
nur einschließen
wenn .....
Hilfe

Untersuchungseinheiten aus Analyse
ausschließen
wenn .....
Hilfe

nur eine Zufallsstichprobe von x %
der Datensätze für Analyse verwenden
Hilfe

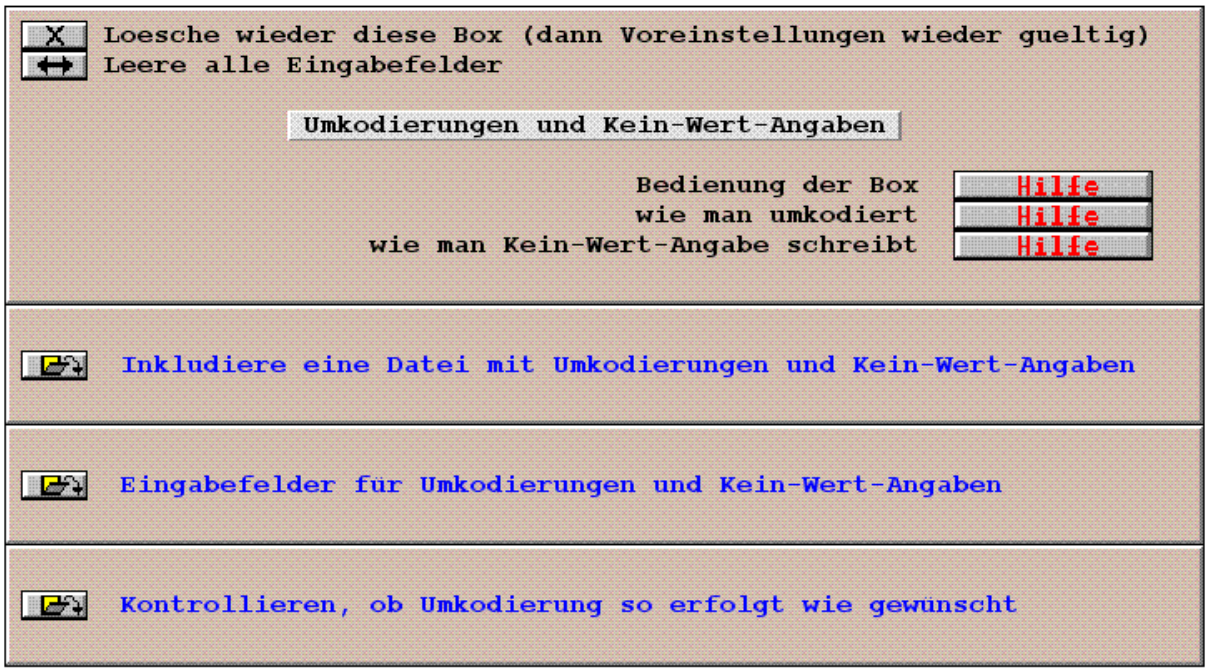
Startwert für Zufallsgenerator
Hilfe

Klicken Sie auf die Hilfeknöpfe und lesen Sie die jeweiligen Hilfen. Eine ausführliche Erläuterung finden Sie im oben angegebenen Handbuch im Abschnitt P0.7.2

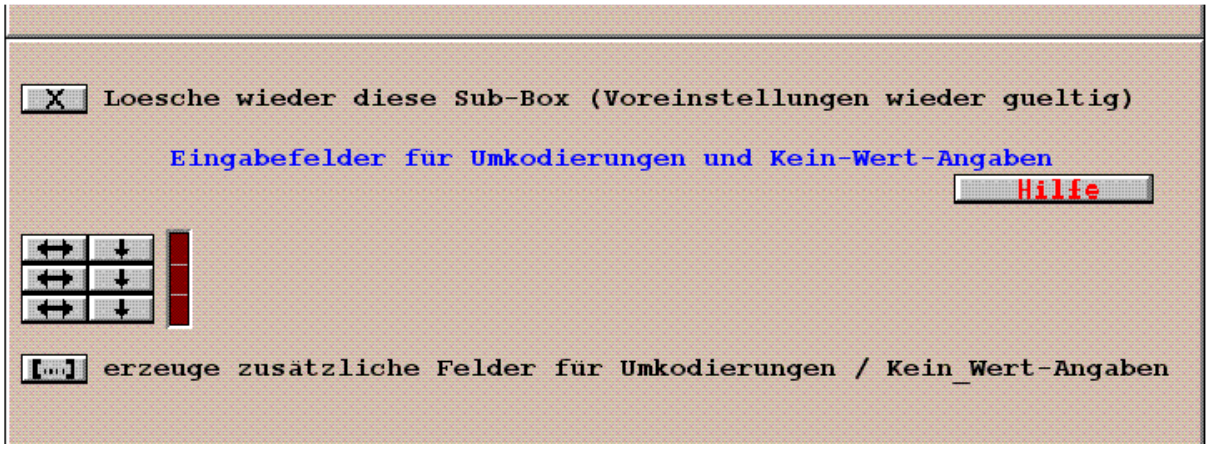
**P14.8.7.3** Umkodierungen

Option: Umkodierungen und Kein-Wert-Angaben

Wird die Optionsbox geöffnet, dann sieht man folgendes



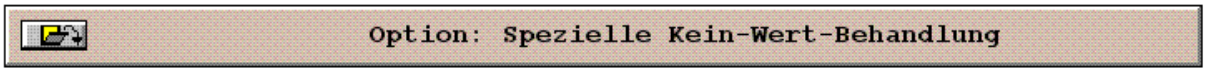

Die Box besteht aus 3 Subboxen. Normalerweise wird man nur die mittlere Subbox öffnen und dort die geplanten Umkodierungen einschreiben




Wie man Umkodierungs-Anweisungen schreibt, erfährt der Benutzer durch Klick auf den Hilfeknopf in der Subbox

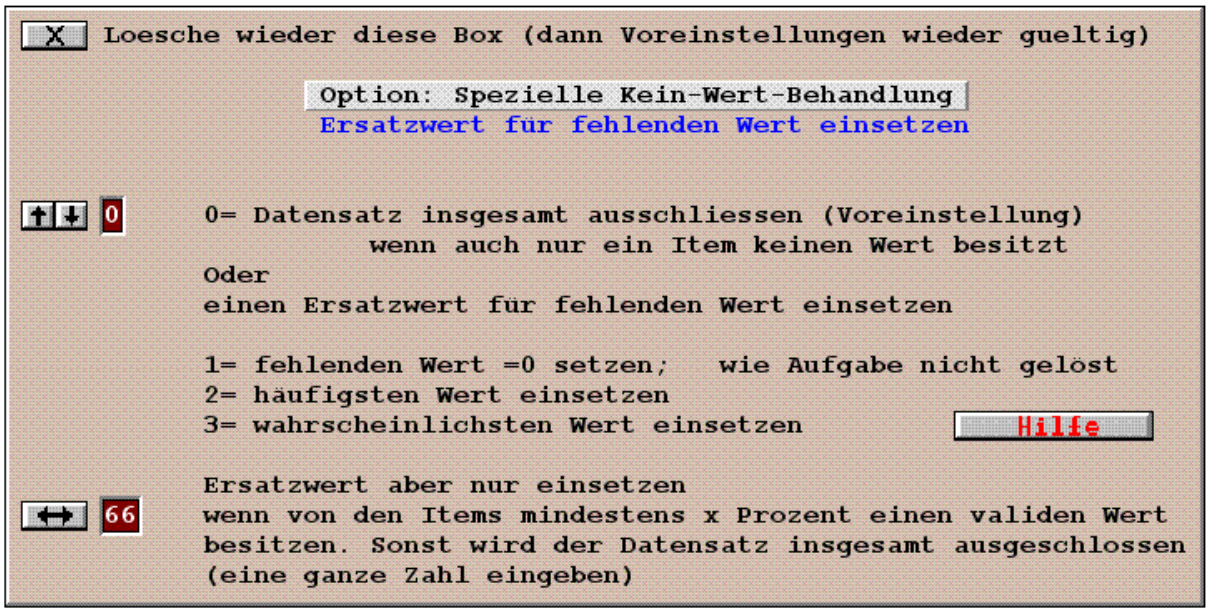
Die komplexen Möglichkeiten, die in Almo zum Umkodieren angeboten werden, sind im oben erwähnten Handbuch, Abschnitt P0.4 ausgeführt.

**P14.8.7.4** Behandlung fehlender Werte




Sind für einige Personen bei einigen Items keine Werte vorhanden, dann treten Probleme auf, die gelöst werden müssen. Standardmäßig schliesst Almo eine Person von der Analyse aus, wenn sie auch nur in einem Item keinen Wert besitzt. Almo bietet jedoch weitere Möglichkeiten an. Nach

dem Öffnen der Box sieht man folgendes




*1. Eingabefeld:*

0= Wenn auch nur einer der Items "Kein-Wert" ist, dann wird der Datensatz insgesamt ausgeschlossen

1= fehlenden Wert =0 setzen; wie nicht gelöst
Hier wird also unterstellt, dass die Person nicht fähig war das Item zu beantworten

2= häufigsten Wert einsetzen
Almo ermittelt, welcher Wert für das betreffende Item am häufigsten auftritt.
Dieser wird als Ersatz für den fehlenden Wert eingesetzt.

3= wahrscheinlichsten Wert als Ersatz einsetzen.

*2. Eingabefeld:*

Der Ersatzwert nach Methode 1 oder 2 oder 3 soll aber nur eingesetzt werden, wenn von den Items mindestens x Prozent einen validen Wert besitzen. Sonst wird der Datensatz insgesamt ausgeschlossen Der Benutzer muss eine ganze Zahl, keine Dezimalzahl eingeben.

***Wahrscheinlichster Wert:***

Der "wahrscheinlichste Wert" wird in folgender Weise ermittelt - an einem Beispiel gezeigt:

Die Variable hat 3 Ausprägungen: (0)leicht, (1)mittel, (2)schwer. Dabei wurden folgende Häufigkeiten ermittelt.

| | Code | Häufigkeit | in % | in % kummuliert |
|---|---|---|---|---|
| leicht | 0 | 250 | 25 | 25 |
| mittel | 1 | 400 | 40 | 65 |
| schwer | 2 | 350 | 35 | 100 |

Dann wird eine gleichverteilte Zufallszahl zwischen 0 und 100 erzeugt.

Liegt sie zwischen

```
 0 und 25, dann wird für den fehlenden Wert 0 eingesetzt
25     65                                   1
65    100                                   2
```

Neben der Einsetzung von Ersatzwerten bietet Almo auch das Imputations-Verfahren an. Der Benutzer muss dazu eines der folgenden Programme rechnen: Prog37m4, Prog45z1, Prog45mz. Die Programme werden erreicht durch Klick auf "Verfahren / Fehlende Werte" Eine ausführliche Beschreibung ist im Handbuch "P45 Data Mining" enthalten.

**P14.8.7.5** Modell-Test

Almo rechnet 3 pauschale Modelltests. "Pauschal" heißt, dass nicht die Brauchbarkeit einzelner Items geprüft wird, sondern nur die gesamte Item-Batterie. Es werden folgende 3 Koeffizienten errechnet:

Likelihood-Ratio-Test
Pearsons Chi-Quadrat
Cressie-Read-Test

*1. Eingabefeld:*

0 = kein Modelltest
1 = Modelltest rechnen
>1 = Modelltest rechnen und zusätzlich Bootstrap-Test

Wird beispielsweise 100 eingegeben, dann werden zusätzlich zum Modelltest 100 Bootstrap-Runden gerechnet.

Der Bootstrap-Test ist dann unerlässlich, wenn Antwortmuster vorhanden sind, die mit weniger als 5 Personen (Erwartungswert) besetzt sind. Das ist fast immer der Fall. Der Benutzer kann sich über die Option "Antwortmuster ausgeben" die empirische und erwartete Besetzungs-Häufigkeit ausgeben lassen.

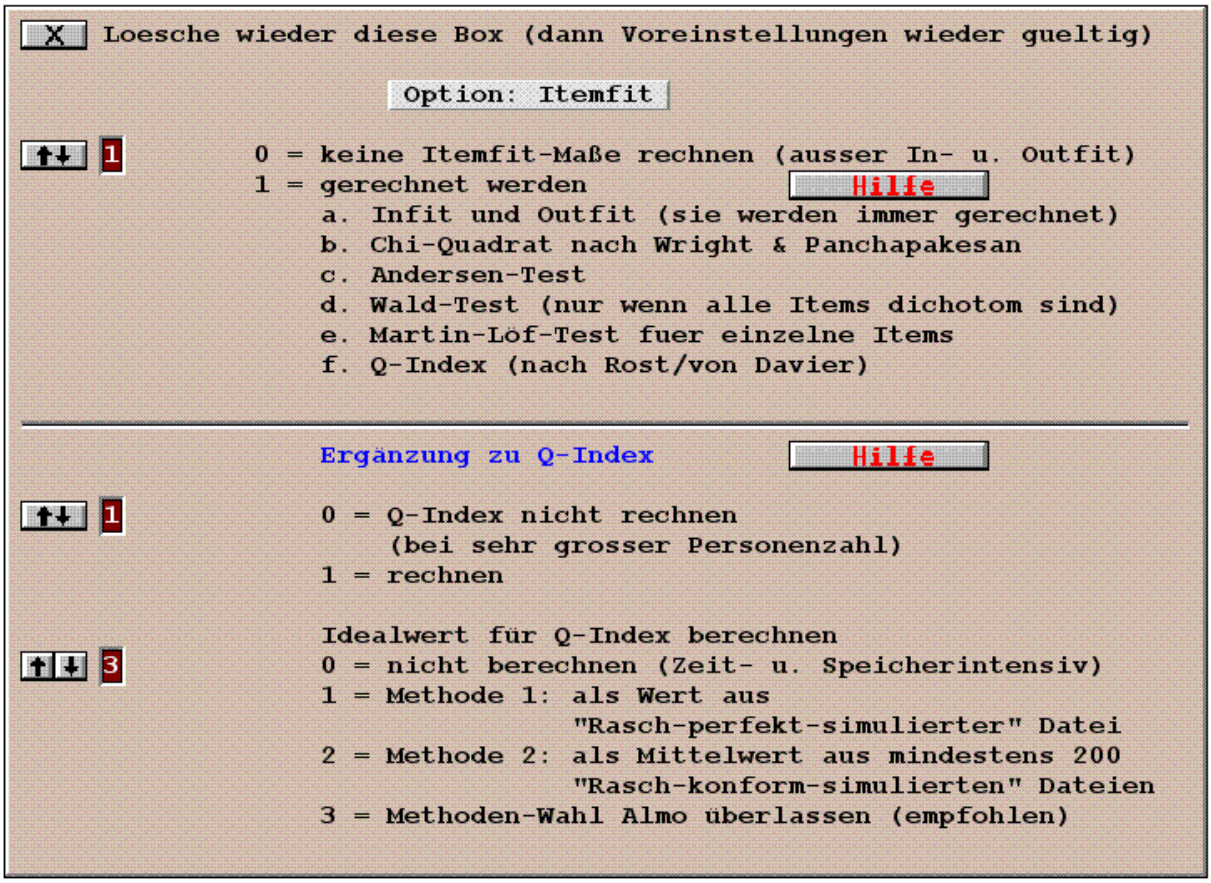
Der Benutzer sollte keine Hemmung davor haben, die Zahl der Bootstrap-Runden auf 500 oder sogar 1000 zu setzen. Das dauert nur wenige Sekunden.

*2. Eingabefeld:*
Der Benutzer kann sich alle oder nur einige Bootstrap-Runden ausgeben lassen. Dabei genügen etwa 20 Runden um einen Eindruck zu bekommen.

*3. Eingabefeld: Zufallszahl.*
Almo besitzt einen eingebauten Zufallsgenerator. Dieser Zufallsgenerator muss durch eine Startzahl in Betrieb gesetzt werden. Die Folge der Zufallszahlen, die Almo dann erzeugt, ist bei jedem Neustart des Programms dieselbe und damit auch das Bootstrap-Ergebnis. Wird die Startzahl geändert, dann wird eine andere Folge von Zufallszahlen erzeugt und damit auch eine andere Folge von Bootstrap-Runden. Verwenden Sie eine ungerade 5- oder 6-stellige Zahl.

**P14.8.7.6 Itemfit**

Geöffnet:




Für den Forscher ist es wichtig zu wissen, welche Items sind gut, welche Items sollten besser aus dem Test eliminiert werden. Man will wissen, welche Items sind "Rasch-konform" und welche

nicht. In der Literatur zum Rasch-Modell werden mehrere Itemfit-Maße angeboten. Bei allen diesen Maßen muss die Frage gestellt werden: Was messen sie eigentlich? Das ist nicht immer so eindeutig. In Almo wird der Begriff „Itemfit“ sehr einfach definiert „als Maß, das informiert, ob ein Item die Voraussetzungen des Rasch-Modells erfüllt oder verletzt“.
Folgende Itemfit-Maße werden erhoben:

a. Infit und Outfit. Dies sind klassische Itemfit-Maße, die in Almo auch gerechnet werden, wenn die Optionsbox nicht geöffnet wird. Sie überprüfen, ob die Trennschärfen der Items (wie im Rasch-Modell gefordert) annähernd gleich sind. Siehe die ausführliche Darstellung unten in Abschnitt P14.8.8.4.1
b. Chi-Quadrat nach Wright & Panchapakesan. Dies ist ein einfacher Chi-Quadrat-Test, der in Almo nur für dichotome Items aktiviert wird. Er überprüft, ob die empirische Lösungswahrscheinlichkeit je Gesamtpunkte ungefähr der vom Rasch-Modell errechneten theoretischen Lösungswahrscheinlichkeit entspricht. Aus der Differenz von empirischen zu theoretischen Lösungswahrscheinlichkeiten wird ein Chi-Quadrat-Wert errechnet.

c. Andersen-Test. Die Personen werden in eine Gruppe mit niedrigem Gesamtpunkt und eine Gruppe mit hohem Gesamtpunkt geteilt. Ihre mittleren Fähigkeiten sind also verschieden. Die beiden Gruppen können auch durch eine Gruppierungsvariable (z.B. Geschlecht: Männer, Frauen) gebildet werden.Die Personen gelten als "Personen-homogen", wenn die verwendeten Items (trotz der unterschiedlichen Fähigkeiten) in beiden Gruppen dieselben Schwierigkeitswerte besitzen.

   Betrachten wir ein Beispiel: Die Person A besitzt eine geringere "Fähigkeit" als die Person B. Das Item i, das die beiden lösen müssen, besitzt eine "Schwierigkeit", die nur ihm anhaftet. "Schwierigkeit" ist eine Eigenschaft des Items. "Fähigkeit" ist eine Eigenschaft der Person. Dies ist das Prinzip der **Separierbarkeit** von Fähigkeit und Schwierigkeit.

   Der Andersen-Test überprüft, ob diese Separierbarkeit gegeben ist.
   Wenn der Andersen-Test negativ ausfällt, dann sind die Personen nicht-homogen - etwa weil sie in unterschiedlichen kulturell-religiösen Denkstrukturen sozialisiert wurden.
   Diese Personen können also nicht durch dasselbe Messinstrument gemessen und verglichen werden. Die Items, die für diese Personen-Inhomogenität verantwortlich sind, werden identifiziert.
   Almo erzeugt zu diesem Zweck ein Koordinatensystem mit den Schwierigkeitswerten der beiden Gruppen als x- und y-Achse. In dieses werden die einzelnen Items bzw. (bei polytomen Items) Itemkategorien als Punkte eingetragen. Die Distanz der Punkte zur Diagonalen ist dann ein Gütekriterium der Items
   Ist ein Itempunkt sehr weit von der Diagonalen entfernt, dann heißt das, dass dieses Item sehr unterschiedliche Schwierigkeitswerte in den beiden Gruppen besitzt. Die Personen werden durch dieses Item nicht in derselben Weise stimuliert; möglicherweise deswegen nicht, weil die Bedingung der Eindimensionalität verletzt ist. Siehe auch nächsten Abschnitt P14.8.7.7
d. Wald-Test. Der Test wird nur gerechnet wenn alle Items dichotom sind. Gemessen wird die "Personen-Homogenität" der Items. Siehe dazu unten

Abschnitt P14. 8.7.7.3

e. Martin-Löf-Test fuer einzelne Items.
Gemessen wird die "Item-Homogenität" der einzelnen Items indem jeweils ein Item allen übrigen gegenüber gestellt wird. Voraussetzung für einen positiven Martin-Löf-Test ist, dass die Items bei den Untersuchungspersonen dieselbe Dimension messen. Siehe dazu auch unten Abschnitt P14.8.7.7.4

f. Q-Index (nach Rost/von Davier). Gemessen wird - grafisch betrachtet - die Lage der logistischen Kurve. Dabei gibt es 2 Extremlagen:
Die Kurve bildet eine horizontale Gerade. Oder die Kurve entspricht einer Guttman-Funktion, also einer einstufigen Treppenkurve.
Der Q-Index drückt den Abstand der realen Kurve von diesen beiden Extremlagen aus. Siehe dazu unten Abschnitt P14.8.8.4.2. Diese anschauliche Interpretation ist natürlich nur bei dichotomen Items möglich. Der Q-Index kann aber auch bei polytomen Items gerechnet werden.

g. Korrelation Itemwert mit Gesamtpunkt und Itemwert mit Fähigkeitswert. Dies ist ein Koeffizient aus der klassischen Testtheorie, der keinen Bezug zum Rasch-Modell besitzt.

Als einen weiteren intuitiven und optischen Itemfit-Test kann der Benutzer die Grafiken der Itemfunktionen benutzen. Er sieht, wie weit die roten Gesamt-Punkte der empirischen Lösungswahrscheinlichkeit vertikal von der Kurve der theoretischen Lösungswahrscheinlichkeit entfernt sind.

Die Frage, welches Itemfit-Maß am besten unbrauchbare Items erkennt, kann nicht beantwortet werden. Alle Itemfit-Masse haben ihre Probleme. Unsere Empfehlung ist, ein Item nur dann nicht zu akzeptieren, wenn es von mehreren Itemfit-Massen verworfen wird und wenn es auch inhaltlich nicht als stimmig empfunden wird. Bei der Itemselektion kann der Forscher sehr frustrierende Erfahrungen machen. Er eliminiert das schlechteste Item in der Erwartung, dass sich der pauschale Modelltest-Wert, z.B. der Cressie-Read-Wert verbessert und erlebt das Gegenteil davon. Es muss offen gelassen werden, ob solche Ergebnisse gegen die Itemfit-Masse oder gegen die pauschalen Modelltests sprechen.

Almo liefert für die 20 dichotomen Mathe-Items aus der Pisa-Studie für Austria folgende Ausgabe (stark gekürzt). Siehe dazu in Almo das Beispielprogramm `Rasch_PisaMathe_A.Alm`

Outfit und Infit

| | | outfit | t *) | | infit | t *) | |
|---|---|---|---|---|---|---|---|
| V1 | M033Q01.MATH.P | 1.16306 | 1.01801 | | 1.11279 | 1.70161 | |
| V2 | M034Q01T.MATH. | 0.91140 | -0.98666 | | 0.97859 | -0.48326 | |
| V3 | M155Q01.MATH.P | 0.83426 | -1.38247 | | 0.90742 | -1.80994 | |
| . | | . | . | | . | . | |
| V7 | M406Q02.MATH.P | 0.63761 | -2.74599 | ? | 0.80425 | -2.99475 | ? |
| V8 | M411Q01.MATH.P | 0.91684 | -0.92820 | | 0.94308 | -1.32705 | |
| V9 | M411Q02.MATH.P | 1.29080 | 3.04023 | ? | 1.18388 | 3.91052 | ? |
| . | | . | . | | . | . | |
| V16 | M564Q02.MATH.P | 1.28155 | 2.98998 | ? | 1.24409 | 5.21052 | ? |
| . | | . | . | | . | . | |
| V19 | M603Q02T.MATH. | 0.79614 | -2.22323 | ? | 0.88150 | -2.51560 | ? |
| V20 | M803Q01T.MATH. | 0.78622 | -1.94620 | | 0.91496 | -1.52844 | |

*) ein t-Wert zwischen -1.96 und +1.96 bedeutet: Die Variable ist "Rasch-konform"
ein nicht-Rasch-konformer Item wird durch ein Fragezeichen ? markiert

Outfit und Infit werden in Abschnitt P14.8.8.4.1 ausführlich erläutert

Chiquadrat-verteilte Prüfgrößen (nach Wright & Panchapakesan)

| Variable | Testwert | Freiheitsgrade | p *) |
|---|---|---|---|
| V1 aaM033Q01.MATH.P | 28.8807 | 20 | 0.0905 |
| V2 aaM034Q01T.MATH. | 16.5624 | 20 | 0.6811 |
| V3 aaM155Q01.MATH.P | 16.2146 | 20 | 0.7031 |
| . | . | . | . |
| V7 aaM406Q02.MATH.P | 26.4274 | 20 | 0.1525 |
| V8 aaM411Q01.MATH.P | 10.5946 | 20 | 0.9560 |
| V9 aaM411Q02.MATH.P | 42.4922 | 20 | 0.0024 ? |
| . | . | . | . |
| V16 aaM564Q02.MATH.P | 40.0546 | 20 | 0.0050 ? |
| . | . | . | . |
| V19 aaM603Q02T.MATH. | 19.4811 | 20 | 0.4909 |
| V20 aaM803Q01T.MATH. | 25.0072 | 20 | 0.2015 |

*) ein p-Wert groesser 0.05 bedeutet: Die empirischen Haeufigkeiten je Gesamtpunkt g der Variablen i sind nicht-signifikant von der theoretischen unterschieden
Die Variable kann als "Rasch-konform" erachtet werden

Es wird der von Wright und Panchapakesan (1969) für das Rasch-Modell entwickelte Test verwendet wird. (vgl. auch Sixtl 1976, S.55). Die Formeln sind folgende

$$y_{ig} = (n_g * p_{ig} - n_g * r_{ig}) \,/\, \sqrt{n_g * r_{ig} * (1 - r_{ig})}$$

$$Chi_i^2 = \sum y_{ig}$$

$$df = g_{max} - g_0$$

$n_g =$ Zahl der Personen mit (demselben) Gesamtpunktwert g

$p_{ig}$ = das ist der empirische Anteil der Personen mit Gesamtpunkt g, die das Item i positiv beantworten.

$r_{ig}$ = theoretische Wahrscheinlichkeit der Personen mit (demselben) Gesamtpunktwert g. Das ist der Ordinaten-Wert (y-Wert) aus der errechneten logistischen Funktion.

df = Freiheitsgrade

$g_{max} =$ maximaler Gesamtpunkt (=Zahl der Items)

$g_0 =$ Zahl der Gesamtpunkte, die nicht besetzt sind. Da für den Gesamtpunkt 0 und den maximalen Gesamtpunkt die empirische Wahrscheinlichkeit $p_{ig}$ notwendigerweise =0 bzw. =1 ist, muss $g_0$ >= 2 sein

Q-Index (nach Rost/von Davier)

| Variable | | Q-Index | Idealwert aus 200 Simulationen | Stand.abweichg | Idealwert minus 2*Stdabwg | Idealwert plus 2*Stdabwg | |
|---|---|---|---|---|---|---|---|
| V1 | M033Q01. | 0.25358 | 0.17505 | 0.0209528941 | 0.13314 | 0.21696 | ? |
| V2 | M034Q01T | 0.16622 | 0.16655 | 0.0190146677 | 0.12852 | 0.20458 | |
| V3 | M155Q01. | 0.14559 | 0.17037 | 0.0181865419 | 0.13400 | 0.20674 | |
| . | | . | . | . | . | | |
| V7 | M406Q02. | 0.09931 | 0.17329 | 0.0240807186 | 0.12513 | 0.22145 | ? |
| . | | . | . | . | . | | |
| V9 | M411Q02. | 0.26727 | 0.16724 | 0.0166062443 | 0.13403 | 0.20045 | ? |
| . | | . | . | . | . | | |
| V16 | M564Q02. | 0.28632 | 0.16802 | 0.0185485578 | 0.13092 | 0.20512 | ? |
| . | | . | . | . | . | | |
| V19 | M603Q02T | 0.12628 | 0.17028 | 0.0181814901 | 0.13392 | 0.20664 | ? |
| V20 | M803Q01T | 0.14261 | 0.17340 | 0.0185570201 | 0.13629 | 0.21052 | |

Der Q-Index wird in Abschnitt P14.8.8.4.2 ausführlich erläutert

Andersen-Test auf Personen-Homogenität: Distanz zur Diagonalen

| Item | Kategorie | Punkt-Nummer | Distanz zu Diagonale | |
|---|---|---|---|---|
| V1 | 1 | 1 | 0.2396 | |
| V2 | 1 | 2 | 0.0269 | |
| V3 | 1 | 3 | 0.5355 | ? |
| V4 | 1 | 4 | 0.1550 | |
| V5 | 1 | 5 | 0.0643 | |
| V6 | 1 | 6 | 0.1108 | |
| V7 | 1 | 7 | 0.1708 | |
| V8 | 1 | 8 | 0.1986 | |
| V9 | 1 | 9 | 0.5233 | ? |
| V10 | 1 | 10 | 0.0205 | |
| V11 | 1 | 11 | 0.2422 | |
| V12 | 1 | 12 | 0.0113 | |
| V13 | 1 | 13 | 0.1233 | |
| V14 | 1 | 14 | 0.1549 | |
| V15 | 1 | 15 | 0.3084 | |
| V16 | 1 | 16 | 0.8675 | ? |
| V17 | 1 | 17 | 0.0560 | |
| V18 | 1 | 18 | 0.0720 | |
| V19 | 1 | 19 | 0.0299 | |
| V20 | 1 | 20 | 0.1327 | |

Der Test wird im folgenden Abschnitt **P14.8.7.7.1** ausführlich erläutert

Test auf Personen-Homogenitaet: **Wald-Test**

```
Die Personen werden entsprechend ihrem Gesamtpunkt
in 2 Gruppen geteilt. Als Teilungskriterium wird verwendet
         der (gerundete) Median: 10


z- und p-Wert je Item und Kategorie
```

```
                                1
                              z       p
   V1  M033Q01.MATH.P    -1.68912  0.09151
   V2  M034Q01T.MATH.    -0.60489  0.54539
   V3  M155Q01.MATH.P     0.42505  0.67089
    .                       .        .
   V7  M406Q02.MATH.P     3.18748  0.00148     ?
    .                       .        .
   V9  M411Q02.MATH.P    -2.42347  0.01555     ?
    .                       .        .
  V16  M564Q02.MATH.P    -5.02406  0.00000     ?
    .                       .        .
  V19  M603Q02T.MATH.     1.45979  0.14466
  V20  M803Q01T.MATH.     2.97032  0.00305     ?

ein p-Wert groesser 0.05 bedeutet, dass das Item zur Personenhomogenitaet beitraegt
Items mit p kleiner 0.05 sollten ausgeschlossen werden
nicht-homogene Items werden durch ein Fragezeichen ? matkiert
```

Martin-Loef-Test auf Item-Homogenitaet fuer einzelne Items

```
Itemgruppe 1 bildet jeweils ein Item, Itemgruppe 2 die uebrigen Items
---------------------------------------------------------------------

Variable    Chi-Quadrat      p*)
   V1         28.2748      0.0605
   V2         18.6125      0.4176
   V3         16.3000      0.5718
    .           .           .
   V7         33.3255      0.0165 ?
    .           .           .
   V9         37.6182      0.0049 ?
    .           .           .
  V16         42.2528      0.0012 ?
    .           .           .
  V19         22.2347      0.2247
  V20         29.3310      0.0467 ?

       Freiheitsgrade= 18

*) ein p-Wert groesser 0.05 bedeutet, dass das Item zur Itemhomogenitaet beitraegt
   nicht-homogene Items werden durch ein Fragezeichen ? matkiert
```

Almo liefert abschliessend eine Tabelle der "problematischen Items". Aus ihr wird ersichtlich, dass V9 und V16 in allen 6 Itemfit-Maßen und V7 in 5 Itemfit-Maßen als problematisch erkannt wurde.

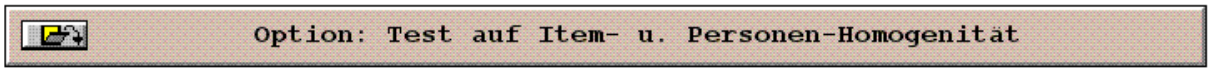
Itemfit-Masse Zusammenfassung: Problematische Items

```
Wald     =Wald-Test auf Personen-Homogenitaet
Wright   =Chi-Quadrat-Test nach Wright & Panchapakesan
Loef     =Martin-Loef-Test auf Homogenitaet von Einzel-Items
Q-Index  =Q-Index nach Rost/v.Davier
Andersen          =Andersen-Test auf Personen-Homogenitaet: Diagonaldistanz
Andersen Gruppvar =Andersen-Test mit Gruppierungsvariable:  Diagonaldistanz

 ? = Item ist problematisch
 - = wurde nicht berechnet

    Variable       Outfit  Infit  Wald  Wright  Loef  Q-Index  Andersen  Andersen
                                                                         Gruppvar
   --------------  ------  -----  ----  ------  ----  -------  --------  --------
```

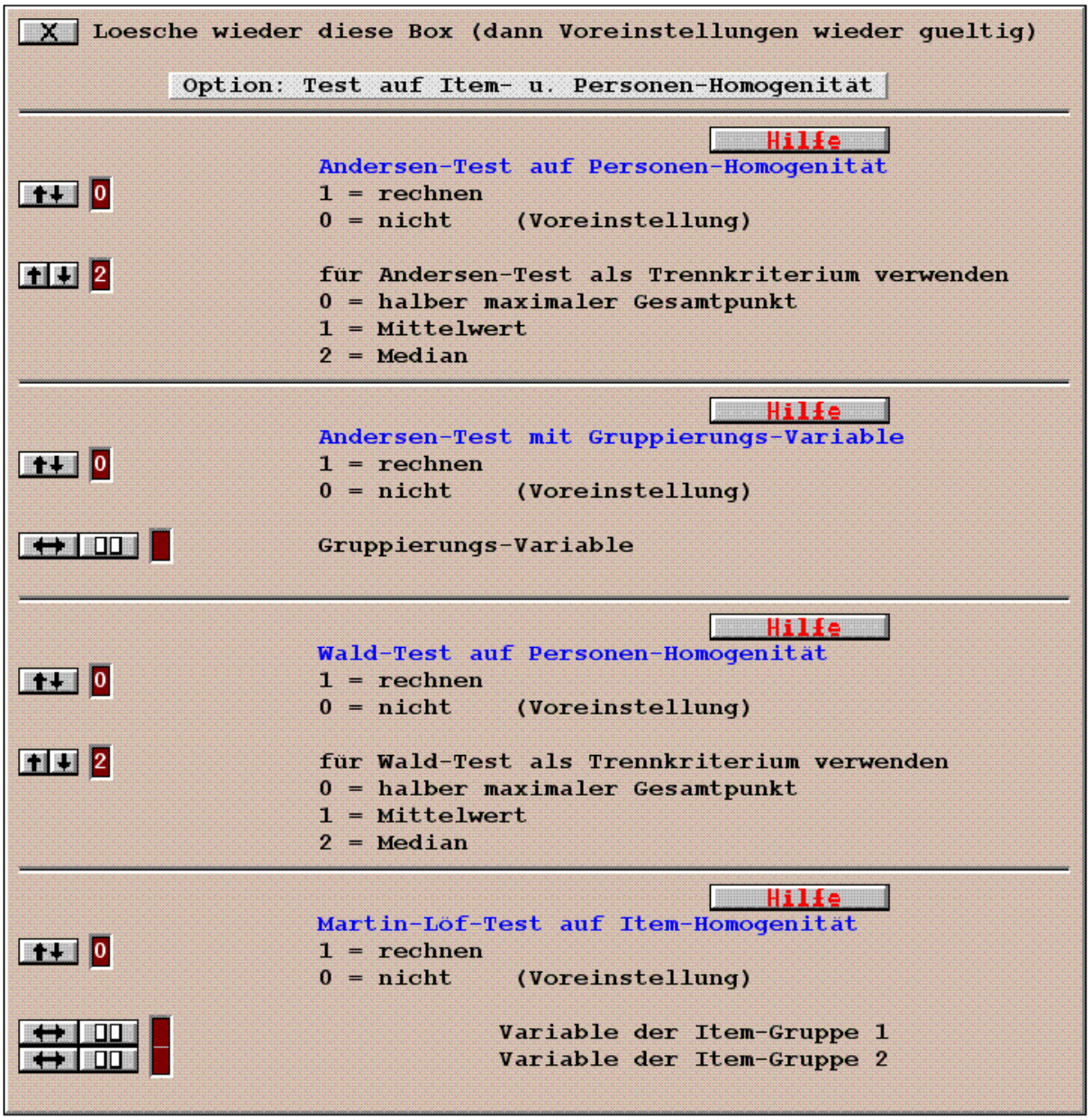
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| V1 | aaM033Q01. | | | | | | ? | | - |
| V2 | aaM034Q01T | | | | | | | | - |
| V3 | aaM155Q01. | | | | | | | ? | - |
| V4 | aaM155Q04T | | ? | | | | ? | | - |
| V5 | aaM192Q01T | | | | | | | | - |
| V6 | aaM406Q01. | | | | | | | | - |
| V7 | aaM406Q02. | ? | ? | ? | | ? | ? | | - |
| V8 | aaM411Q01. | | | | | | | | - |
| V9 | aaM411Q02. | ? | ? | ? | ? | ? | ? | ? | - |
| V10 | aaM423Q01. | ? | | | ? | ? | ? | | - |
| V11 | aaM442Q02. | | ? | ? | ? | ? | | | - |
| V12 | aaM474Q01. | | ? | ? | ? | ? | ? | | - |
| V13 | aaM496Q01T | | | | | | | | - |
| V14 | aaM496Q02. | | | | | | | | - |
| V15 | aaM564Q01. | | ? | | | | ? | | - |
| V16 | aaM564Q02. | ? | ? | ? | ? | ? | ? | ? | - |
| V17 | aaM571Q01. | | | | | | | | - |
| V18 | aaM603Q01T | | | | | | | | - |
| V19 | aaM603Q02T | ? | ? | | | | ? | | - |
| V20 | aaM803Q01T | | | ? | | ? | | | - |

Item V9 und V16 werden von allen Itemfit-Maßen als problematisch eingestuft, die Items V7 und 12 von 5 Maßen. Werden diese 4 Items herausgenommen, dann erweisen sich in einer neuen Analyse die Items V10 und 15 mit je 6 Fragezeichen als problematisch. Unsere Empfehlung ist, die Ergebnisse aus den Itemfit-Maßen zu benützen, um problematische Items inhaltlich zu überdenken.

### P14.8.7.7 Test auf Personen-Homogenität




Geöffnet:




Zwei Gruppen von Personen sind im Sinne des Rasch-Models "homogen", wenn die Items, die Ihnen vorgelegt wurden, bei ihnen dieselben Schwierigkeitswerte besitzen. Siehe auch die ausführliche Erläuterung oben in Abschnitt P14.8.7.6, Punkt c

Der Andersen-Test ist in Almo in 2 Varianten enthalten.

In der 1. Variante werden die Personen in eine Gruppe mit niedrigem Gesamtpunkt und eine Gruppe mit hohem Gesamtpunkt geteilt. Ihre mittleren Fähigkeiten sind also verschieden. Die Personen gelten als "Rasch-homogen", wenn die verwendeten Items (trotz der unterschiedlichen Fähigkeiten) in beiden Gruppen dieselben Schwierigkeitswerte besitzen. Das mag nicht besonders einsichtig sein.

Betrachten wir ein Beispiel: Die Person A besitzt eine geringere "Fähigkeit" als die Person B. Das Item i, das die beiden lösen müssen, besitzt eine "Schwierigkeit", die nur ihm anhaftet. "Schwierigkeit" ist eine Eigenschaft des Items. "Fähigkeit" ist eine Eigenschaft der Person. Dies ist das Prinzip der **Separierbarkeit** von Fähigkeit und Schwierigkeit. Der Andersen-Test überprüft, ob diese Separierbarkeit gegeben ist.

In der 2. Variante des Andersen-Tests werden die Personen entsprechend einer "Gruppierungs-variablen" in 2 Gruppen geteilt - etwa nach Geschlecht.

**P14.8.7.7.1 Andersen-Test**

Beim Andersen-Test werden die Personen in eine Gruppe mit niedrigem Gesamtpunkt und eine Gruppe mit hohem Gesamtpunkt geteilt. In Almo werden 3 Teilungskriterien angeboten:

*2. Eingabefeld:*

0 = der halbe maximale Gesamtpunkt. Beispiel: Die Gesamtpunkte sind 0, 1, 2, 3, 4, 5. Die Mitte ist 2.5, abgerundet auf 2. Die "untere" Gruppe besteht dann aus den Personen, die 0, 1, 2 Gesamtpunkte erreicht haben, die "obere" Gruppe aus denen die 3, 4, 5, Punkte erreicht haben. Der Nachteil des Verfahrens ist, dass die Personenzahl in den beiden Gruppen stark unterschiedlich sein kann.

1 = der auf Ganzzahligkeit gerundete Mittelwert - berechnet aus allen Personen

2 = der auf Ganzzahligkeit gerundete Median - berechnet aus allen Personen. In diesem Fall sind die beiden Gruppen (fast) gleich groß

Für beide Gruppen werden nun getrennt die Schwierigkeitsparameter aller Items ermittel. Sind sie gleich, dann dann sind die Personen "homogen".

Als Beispiel für die Almo-Ausgabe werden die 20 Pisa-Mathe-Items für Austria verwendet. Die Ausgabe ist stark gekürzt. Siehe in Almo das Beispielprogramm `Rasch_PisaMathe_A.Alm`

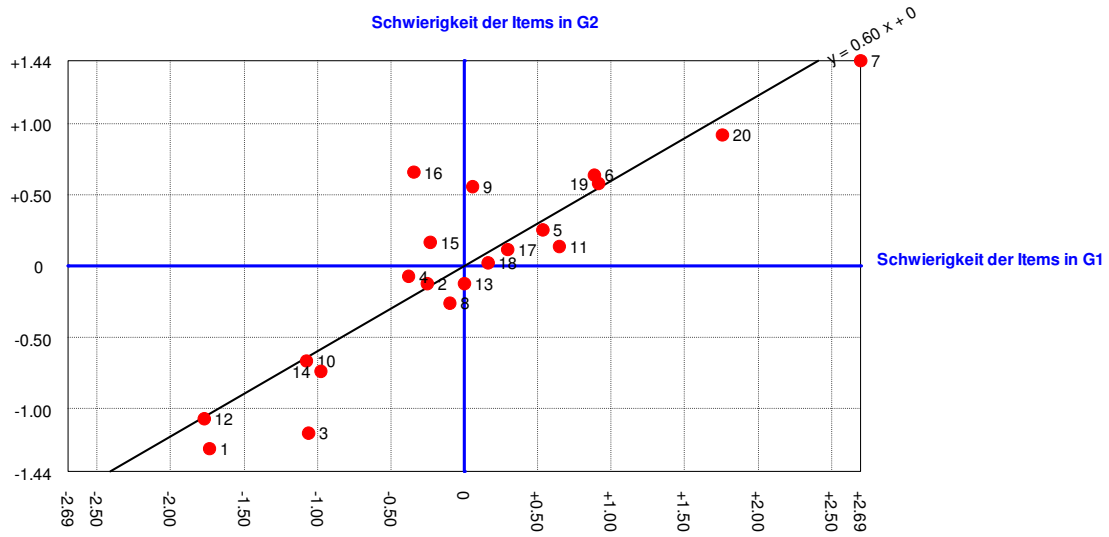
```
Haeufikeiten je Variable und Kategorie in Gruppe 1 (G1) und Gruppe 2 (G2)
                 0            1
            -------      -------
    V1  G1      80          169
        G2      20          198

    V2  G1     163           86
        G2      52          166

    V3  G1     117          132
        G2      22          196
         .       .            .
         .       .            .
   V19  G1     213           36
        G2      83          135

   V20  G1     232           17
        G2     100          118
```

```
Schwierigkeitsparameter je Variable und Kategorie in Gruppe 1 (G1) und Gruppe
2 (G2)
              0            1
           -------      -------
    V1  G1     -        -1.73282
        G2     -        -1.27676

    V2  G1     -        -0.25498
        G2     -        -0.12576

    V3  G1     -        -1.06072
        G2     -        -1.17038
         .     .          .
         .     .          .
         .     .          .
   V19  G1     -         0.91583
        G2     -         0.57802

   V20  G1     -         1.75922
        G2     -         0.92023

Chi-Quadrat-Test auf Gleichheit der Schwierigkeitsparameter der beiden Gruppen
Chi-Quadrat=  84.1732  df=19  p= 0.0000  *)

Ein p-Wert groesser 0.05 bedeutet, dass die Schwierigkeit der Items
fuer beide Personengruppen gleich ist.
```

Dies ist nicht der Fall. Für die Pisa-Mathe-Items für Austria kann keine Personenhomogenität festgestellt werden.
Beachte: Die Schwierigkeitswerte sind je Gruppe nullsummen-normiert, so dass sie auch verglichen werden können. Bei perfekter Personenhomogenität sind sie gleich plus/minus eine Konstante.

Die Ergebnisse des Andersen-Tests lassen sich auch grafisch sehr eindrucksvoll darstellen.

Streudiagramm



```
*** Die Punkte der Items sollten auf der Diagonalen liegen
    Die Distanz zur Diagonalen ist ein Indikator für die „Güte" des Items

Steigung der Diagonalen (=Regressionsgerade): 0.599 (=31 Grad)
```

Almo liefert zur Grafik folgende Tabelle der Distanzen

```
Die roten Punkte im Streudiagramm entsprechen folgenden Items bzw. Kategorien
(Problematische Items werden durch ein Fragezeichen ? markiert)
```

| Item | Kategorie | Punkt-Nummer | Distanz zu Diagonale |
|---|---|---|---|
| V1 | 1 | 1 | 0.2396 |
| V2 | 1 | 2 | 0.0269 |
| V3 | 1 | 3 | 0.5355 ? |
| V4 | 1 | 4 | 0.1550 |
| V5 | 1 | 5 | 0.0643 |
| V6 | 1 | 6 | 0.1108 |
| V7 | 1 | 7 | 0.1708 |
| V8 | 1 | 8 | 0.1986 |
| V9 | 1 | 9 | 0.5233 ? |
| V10 | 1 | 10 | 0.0205 |
| V11 | 1 | 11 | 0.2422 |
| V12 | 1 | 12 | 0.0113 |
| V13 | 1 | 13 | 0.1233 |
| V14 | 1 | 14 | 0.1549 |
| V15 | 1 | 15 | 0.3084 |
| V16 | 1 | 16 | 0.8675 ? |
| V17 | 1 | 17 | 0.0560 |
| V18 | 1 | 18 | 0.0720 |
| V19 | 1 | 19 | 0.0299 |
| V20 | 1 | 20 | 0.1327 |

Die Distanzen können auch für die einzelnen Kategorien eines polytomen Items berechnet werden. Jeder Punkt im Streudiagramm repräsentiert eine Itemkategorie.

Auffällig im Streudiagramm ist, dass die beiden Koordinatenachsen verschieden lang sind. Die Items 18, 17, 5, 11, und besonders die schweren Items 19, 6, 20, 7 sind für die untere Gruppe G1 (die "Schwachen") schwieriger als für die obere Gruppe G2 (die "Starken"). Das ist psychologisch verständlich, aber nicht "Rasch-konform".

Andererseits sind die Items 2, 4, 10, 14 und besonder die leichten Items 1 und 12 für die "schwachen" leichter als für die "Starken". Ein überraschendes Ergebnis.
Die Achse der "Schwachen" ist also stärker gespreizt als die Achse der "Starken". Für die "Starken" ist sozusagen der Schwierigkeitsbereich der Items enger.

Diese „Spreizung" bei den „Schwachen" ist ein Ergebnis das man sehr häufig bei empirischen Daten antreffen kann, das aber psychologisch verständlich ist.

Bei "Rasch-konformen" Daten darf jedoch ein solches Phänomen nicht auftreten.
Das Rasch-Modell ist hier gegenüber realen Daten zu rigoros.

Nachfolgend wird diese Notation verwendet:

```
m      = Zahl der Items
s1     = Schwierigkeitwert der Itemkatorien in Gruppe 1 (x-Achse)
s2     = Schwierigkeitwert der Itemkatorien in Gruppe 2 (y-Achse)
s2_max = absolut grösster Schwierigkeitswert aus s2
d      = Distanz eines Itempunktes von der Diagonalen
reg    = Koeffizient der Regression s2 auf s1
```

Die entscheidende Frage ist, wenn die Koordinatenachsen in der Regel verschieden lang sind - wie soll dann die Lage der Diagonalen bestimmt werden. Auf diese bezogen sollen ja die Distanzen der Items berechnet werden.

In Almo wird als Diagonale die durch die Punkte verlaufende Regressiongerade verwendet. Der Regressionskoeffizient 'reg' ist hier sehr einfach zu bestimmen.

```
reg = Summe(s1*s2) / Summe(s1*s1)  für alle Items 1 bis m
```

Da s1 und s2 jeweils auf Summe 0 normiert sind gibt es keine Konstante.

Die Regressionsgerade ist dann

```
s2 = reg*s1
```

Die Regressionsgerade als Diagonale hat folgende (positive) Eigenschaften:

1. Die Diagonale verläuft immer durch den Ursprung des Koordinatensystems, da s1 und s2 nullsummennormiert sind.
2. Sie verläuft so, dass die (quadrierten) Distanzen der Punkte zur Diagonalen in Summe ein Minimum sind. Es wird also sozusagen "das Beste aus den Items herausgeholt"
3. Sind s1 und s2 identisch, dann hat die Diagonale eine Steigung von 1 (45 Grad) Weicht die tatsächliche Steigung von 45 Grad erheblich ab, dann besteht keine Personenhomogenität. Der Andersen Chi-Quadrat-Test wird dann negativ ausfallen. Wenn die Steigung der Diagonalen nahe 45 Grad ist, dann kann der Andersen Chi-Quadrat-Test trotzdem negativ ausfallen. Dies geschieht dann, wenn die Itempunkte zu weit um die Diagonale streuen. Der Andersen-Test reagiert also (1) auf die Abweichung von der 45 Grad-Steigung und (2) auf das Maß der Streung der Itempunkte um die Diagonale. In Almo wird, wenn die Distanz als Gütemaß des Items interpretiert werden soll, Punkt (1) nicht berücksichtigt.

In Almo wird nun (unabhängig von der Steigung) die Streuung der Itempunkte um die Diagonalen untersucht. Der Andersen-Test wird so zur "Itemfit-Prüfung" benutzt. Die schlechten Items sollen relativ zu den anderen identifiziert werden. Almo akzeptiert also eine von 1 abweichende Steigung der Diagonalen und auch einen ungünstigen Andersen-Chi-Quadrat-Wert. Es geht darum, die Distanz der Items zur Diagonalen zu errechnen und diese Distanz als ein Maß der Güte der Items zu interpretieren.

Die Distanz eines Itempunktes von der Diagonalen ist

```
d = abs( s2-reg*s1 )
```

In der Sprache der Regressionsanalyse wäre dies das "Residuum" der Itemkategorie. Gemessen wird die Distanz d in den Maßeinheiten der y-Achse also der Schwierigkeitswerte der Gruppe G2.

Die Distanz d wird dann verglichen mit s2_max, dem absolut grössten Schwierigkeitswert der Gruppe G2. Almo markiert in der Ergebnisliste ein Item (durch ein Fragezeichen) als "problematisch", wenn es weiter als 0.25*s2_max von der Diagonale entfernt ist.

```
d > 0.25*s2_max   problematisch !!
```

Almo gibt die Tabelle der Distanzen der Itempunkte von der Diagonalen nochmals in erweiterter Form aus. Die Tabelle enthält dann folgende Informationen;

```
A = Distanz des Itempunktes zur Diagonalen
    Items, die 25 % vom absolut groessten Schwierigkeitswert
    von G2 ueberschreiten, sind durch ein Fragezeichen markiert

B = mittlere Distanz aus 500 Rasch-konform simulierten Zufallsdateien zu je
    N Personen

C = B plus 2 mal die Standardabweichung aus den 200 Distanzen
    Itemkategorien, die C ueberschreiten sind durch ein Fragezeichen markiert

D = Zahl der Faelle (von den 500), bei denen der Distanzwert aus der
    Simulation groesser (also schlechter) war als bei den Untersuchungsdaten

E = p ist dann der Anteilswert dieser Faelle. p muss groesser sein als 0.05
    Itemkategorien, die p=0.05 unterschreiten sind durch ein Fragezeichen
    markiert
```

| Item | Kategorie | Punkt Nummer | A<br>Distanz zu Diagonale | B<br>mittlere Distanz aus 500 Simulationen | C<br>mittlere Distanz +2*Standabwg aus 500 Simulationen | D<br>Distanz aus 500 Simulationen ist groesser Faelle | E<br>p | |
|---|---|---|---|---|---|---|---|---|
| V1 | 1 | 1 | 0.2396 | 0.2081 | 0.5543 | 164 | 0.3280 | |
| V2 | 1 | 2 | 0.0269 | 0.1791 | 0.4373 | 462 | 0.9240 | |
| V3 | 1 | 3 | 0.5355 ? | 0.1981 | 0.5008 ? | 13 | 0.0260 | ? |
| V4 | 1 | 4 | 0.1550 | 0.1705 | 0.4427 | 240 | 0.4800 | |
| V5 | 1 | 5 | 0.0643 | 0.1734 | 0.4243 | 390 | 0.7800 | |
| V6 | 1 | 6 | 0.1108 | 0.1779 | 0.4534 | 312 | 0.6240 | |
| V7 | 1 | 7 | 0.1708 | 0.1868 | 0.4541 | 246 | 0.4920 | |
| V8 | 1 | 8 | 0.1986 | 0.1570 | 0.3988 | 151 | 0.3020 | |
| V9 | 1 | 9 | 0.5233 ? | 0.1610 | 0.4166 ? | 6 | 0.0120 | ? |
| V10 | 1 | 10 | 0.0205 | 0.1970 | 0.4896 | 467 | 0.9340 | |
| V11 | 1 | 11 | 0.2422 | 0.1671 | 0.4306 | 130 | 0.2600 | |
| V12 | 1 | 12 | 0.0113 | 0.2094 | 0.5307 | 492 | 0.9840 | |
| V13 | 1 | 13 | 0.1233 | 0.1762 | 0.4511 | 273 | 0.5460 | |
| V14 | 1 | 14 | 0.1549 | 0.1787 | 0.4552 | 244 | 0.4880 | |
| V15 | 1 | 15 | 0.3084 | 0.1643 | 0.4088 | 65 | 0.1300 | |
| V16 | 1 | 16 | 0.8675 ? | 0.1650 | 0.4141 ? | 0 | 0.0000 | ? |
| V17 | 1 | 17 | 0.0560 | 0.1586 | 0.4071 | 385 | 0.7700 | |
| V18 | 1 | 18 | 0.0720 | 0.1781 | 0.4311 | 374 | 0.7480 | |
| V19 | 1 | 19 | 0.0299 | 0.1690 | 0.4252 | 449 | 0.8980 | |
| V20 | 1 | 20 | 0.1327 | 0.1846 | 0.4527 | 285 | 0.5700 | |

```
Itemdistanzen groesser 0.3601 sind problematisch
sie werden durch Fragezeichen markiert
```

```
alle Distanzen sind in Einheiten der y-Achse
d.h. in Einheiten der Schwierigkeitswerte der Gruppe G2 gemessen
```

Almo simuliert 500 "Rasch-konforme" Zufallsdateien zu je N Personen.

Wie beim bootstrapping simuliert Almo mit Zufallszahlen eine Rasch-konforme Datei - wobei aus den Untersuchungsdaten folgende Parameter übernommen werden:

m Zahl der Items,
N Zahl der Personen,
s Schwierigkeitsparameter,
f Fähigkeitsparameter,
G Personenzahl je Gesamtpunkt.

Als Trennschärfeparameter (der die Steilheit der Kurve bestimmt) wird 1 eingesetzt.

Siehe das C-Programm "simord" in den Algorithmen_in_C im Modul up_algorith11.c oder das entsprechende R-Programm in „Rasch_in_R“. Es wird also eine N*m-Datei erzeugt, die der empirischen Datei äquivalent ist – mit dem entscheidenden Unterschied, dass die simulierten Daten "Rasch-konform" sind. Für diese Datei werden je Item (bzw. je Itemkategorie) die Distanzen zur Diagonalen errechnet. Das wird 500 mal wiederholt, d.h. es werden 500 Dateien zu je N Personen erzeugt. So entstehen für jede der Itemkategorien 500 Distanz-Werte. Daraus wird dann für jede Itemkategorie der Mittelwert der Distanzen und die Standardabweichung ermittelt. Die mittlere Distanz wird in Spalte B ausgegeben.

Diese simulierten mittleren Distanzen werden mit den tatsächlichen Distanzen verglichen. Dafür gibt es 2 Möglichkeiten.

1. Spalte C
Es wird untersucht, ob die tatsächliche Distanz kleiner ist als die mittlere simulierte Distanz plus 2 Standardabweichungen. Ist dies nicht der Fall, dann wird die betreffende Itemkategorie als "problematisch" durch ein Fragezeichen markiert

2. Spalte D und E
Es wird ausgezählt, wie oft die simulierte mittlere Distanz grösser ist als die tatsächliche. Beispiel: Bei Item V2 ist sie in 462 Fällen (von 500) grösser - also mit einem Anteilswert von p=0.9240. Bei Item V3 ist sie nur 13 mal (von 500) grösser - also mit einem Anteilswert von p=0.0260. Item V2 ist brauchbar. Item V3 ist "problematisch" und wird durch ein Fragezeichen markiert. Items mit einem p-Wert kleinergleich 0.05 werden als "problematisch" markiert.

Wurde die Optionsbox "Itemfit" geöffnet, dann gibt Almo in der Ergebnisdatei unter folgender Überschrift aus

```
Itemfit-Masse Zusammenfassung: Problematische Items
```

Dort werden in der Spalte "Andersen" die Fragezeichen aus obiger Spalte C eingetragen.

#### P14.8.7.7.2 Andersen-Test mit Gruppierungsvariable

Es werden entsprechend einer dichotomen Gruppierungsvariablen 2 Personengruppen gebildet.

*4. Eingabefeld:*
Die Gruppierungsvariable könnte beispielsweise das Geschlecht sein. Es muss eine Variablennummer (z.B. V48) oder ein Variablennamen eingegeben werden. Die Gruppierungsvariable darf nur 2 Ausprägungen besitzen. Die beiden Zahlenwerte dieser Variablen müssen aufeinander folgen, z.B. 3, 4.

Der Andersen-Test überprüft, ob die Items für beide Gruppen dieselben Schwierigkeitsparameter besitzen.

Ein Beispiel für diese Variante des Andersen-Tests ist folgendes: An der Universität Wien wurde ein umfangreicher Aufnahme-Test für Medizin-Studenten durchgeführt. Bei der Auswertung ergab es sich, dass Frauen durchschnittlich schlechter abgeschnitten haben als Männer. Wäre ein Rasch-Modell mit einem Andersen-Test mit Gruppierungsvariable gerechnet worden, dann hätte entschieden werden können, ob

> die Schwierigkeitswerte der verwendeten Items für die Gruppe der Frauen und der Männer gleich waren - die Fähigkeitswerte der Frauen im Durchschnitt jedoch niedriger waren. Das wäre dann der Fall der "Homogenität" der Personen.
>
> Oder ob
>
> die Schwierigkeitswerte der verwendeten Items für die beiden Gruppen ungleich war - der Test also nicht „Rasch-konform“ war.

Als Beispiel betrachten wir wieder die Pisa-Mathe-Items. Die Ausgabe ist stark gekürzt. Als Gruppierungsvarible wird das Geschlecht verwendet. Für Mädchen und Buben werden getrennt die Schwierigkeitswerte der 20 Items ermittelt.

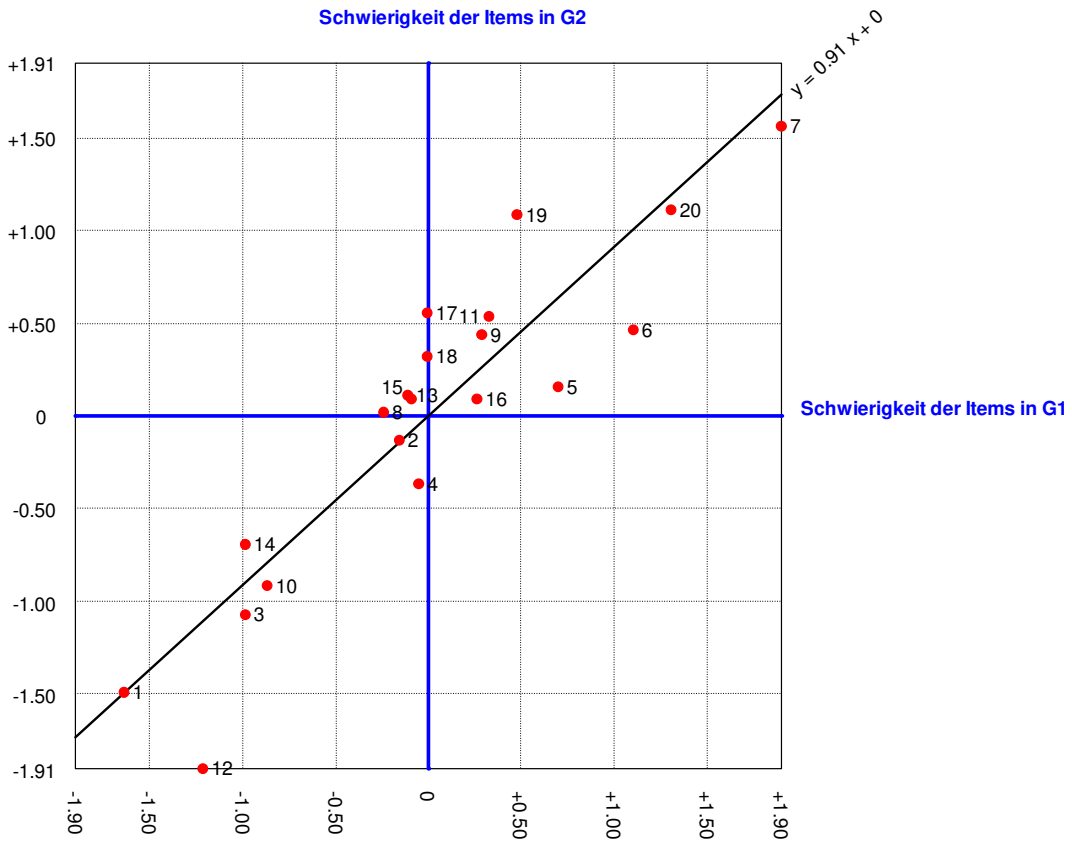
```
Test auf Gleichheit der Item-Schwierigkeiten fuer 2 Personengruppen
Gruppierungsvariable: V21 Geschlecht: weibl (Gruppe1),  maennl (Gruppe2)

Schwierigkeitsparameter je Variable und Kategorie in Gruppe 1 (G1) und Gruppe
2 (G2)
              0            1
           -------      -------
    V1  G1     -       -1.63337
        G2     -       -1.49020

    V2  G1     -       -0.15860
        G2     -       -0.13559

    V3  G1     -       -0.98439
        G2     -       -1.07483
         .                 .
         .                 .
         .                 .

Chi-Quadrat-Test auf Gleichheit der Schwierigkeitsparameter der beiden Gruppen
Chi-Quadrat=  52.8459  df=19  p= 0.0001
```

Der p-Wert ist wieder kleiner alsr 0.05. Das bedeutet, dass die Schwierigkeit der Items für Mädchen und Buben in Österreich verschieden ist

Die Ergebnisse des Andersen-Tests mit Gruppierungsvariable werden auch grafisch dargestellt. Almo liefert folgende Ausgabe.

Streudiagramm der Schwierigkeitswerte der Items in Gruppe G1 (Mädchen) und G2 (Buben)

Streudiagramm




```
*** Die Punkte der Items sollten auf der Diagonalen liegen
    Die Distanz zur Diagonalen ist ein Indikator für die „Güte" des Items

Steigung der Diagonalen (=Regressionsgerade): 0.914 (=42 Grad)

Die roten Punkte im Streudiagramm entsprechen folgenden Items bzw. Kategorien
(Problematische Items werden durch ein Fragezeichen ? markiert)
```

| Item | Kategorie | Punkt-Nummer | Distanz zu Diagonale |
|---|---|---|---|
| V1 | 1 | 1 | 0.1074 |
| V2 | 1 | 2 | 0.0169 |
| V3 | 1 | 3 | 0.0599 |
| V4 | 1 | 4 | 0.2212 |

```
   V5       1           5          0.3820  ?
   V6       1           6          0.4522  ?
    .       .           .           .
    .       .           .           .
    .       .           .           .
  V12       1          12          0.4864  ?
  V13       1          13          0.1651
  V14       1          14          0.2054
  V15       1          15          0.1339
  V16       1          16          0.1214
  V17       1          17          0.4032  ?
  V18       1          18          0.2370
  V19       1          19          0.4257  ?
  V20       1          20          0.1412
```

In der empirischen Forschung sollte der Andersen-Test für eine Gruppierungsvariable nicht unkritisch dazu verwendet werden, „problematische“ Items zu identifizieren. Nicht jedes Gruppierungsmerkmal eignet sich dazu. Allerdings gilt rigoros: „Rasch-konforme“ Items müssen für zwei verschiedene Gruppen den gleichen Schwierigkeitswert besitzen, egal welches Gruppierungsmerkmal verwendet wird.

### P14.8.7.7.3 Wald-Test auf Personen-Homogenität

Der Waldtest verläuft genauso wie der oben beschriebene Andersen-Test. Dabei besitzt er jedoch einen bedeutsamen Vorteil, aber leider auch einen Nachteil. Sein Vorteil: Er überprüft für jedes einzelne Item ob es die Homogenität der Personen stört (und somit ausgeschlossen werden sollte). Sein Nachteil: Er ist nur sinnvoll, wenn alle Items dichotom sind.

Zuerst werden wie beim Andersen-Test die Schwierigkeitsparameter für die beiden Personengruppen ausgegeben. Dann jedoch folgt die Ausgabe der p-Werte je Item

```
z- und p-Wert je Item und Kategorie
                                1
                             z        p
   V1  M033Q01.MATH.P  -1.68912  0.09151
   V2  M034Q01T.MATH.  -0.60489  0.54539
   V3  M155Q01.MATH.P   0.42505  0.67089
   V4  M155Q04T.MATH.  -1.46254  0.14391
   V5  M192Q01T.MATH.   1.26718  0.20539
   V6  M406Q01.MATH.P   1.06168  0.28864
   V7  M406Q02.MATH.P   3.18748  0.00148
   V8  M411Q01.MATH.P   0.71768  0.47312
   V9  M411Q02.MATH.P  -2.42347  0.01555
  V10  M423Q01.MATH.P  -1.79411  0.07310
  V11  M442Q02.MATH.P   2.21447  0.02702
  V12  M474Q01.MATH.P  -2.72757  0.00649
  V13  M496Q01T.MATH.   0.56109  0.57486
  V14  M496Q02.MATH.P  -1.01493  0.31039
  V15  M564Q01.MATH.P  -1.94393  0.05218
  V16  M564Q02.MATH.P  -5.02406  0.00000
  V17  M571Q01.MATH.P   0.80355  0.42185
  V18  M603Q01T.MATH.   0.63384  0.52633
  V19  M603Q02T.MATH.   1.45979  0.14466
  V20  M803Q01T.MATH.   2.97032  0.00305
```

```
ein p-Wert groesser 0.05 bedeutet, dass das Item zur Personenhomogenitaet beitraegt
```

Ginge es nach dem Wald-Test, dann müssten eine ganze Reihe von Items ausgeschlossen werden

#### P14.8.7.7.4 Martin-Löf-Test auf Item-Homogenität

Items sind homogen, wenn sie an den Untersuchungs-Personen dieselbe Fähigkeit (allgemein: dieselbe Dimension) messen. Die Items werden in 2 Gruppen geteilt. Die beiden Gruppen von Items sind im Sinne des Rasch-Models homogen, wenn dieselben Personen aus der Analyse für Itemgruppe 1 und für Itemgruppe 2 dieselben Fähigkeitswerte erhalten.

Im letzten Eingabefeld werden die Analysevariable in 2 Gruppen getrennt. Jedes Item muss in einer der beiden Gruppen enthalten sein.

Es sollten verschiedene Gruppierungen durchprobiert werden. Im nachfolgenden Beispiel wurden für die Pisa-Mathe-Items zwei Gruppen gebildet

```
Die Items wurden in 2 Gruppen geteilt
Item-Gruppe 1: V1 V2 V3 V4 V5 V6 V7 V8 V9 V10
Item-Gruppe 2: V11 V12 V13 V14 V15 V16 V17 V18 V19 V20

Zahl der Personen, die in der Gruppe 1 x Gesamtpunkte
und in der Gruppe 2 y Gesamtpunkte erzielt haben
```

| | | Items Gruppe 2 Gesamtpunkte | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Items Gruppe 1 Ges.pkte. | 0 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 8 | 8 | 4 | 4 | 2 | 0 | 1 | 0 | 0 | 0 |
| | 2 | 2 | 9 | 10 | 8 | 9 | 2 | 2 | 0 | 0 | 0 | 0 |
| | 3 | 4 | 7 | 15 | 8 | 5 | 7 | 6 | 3 | 0 | 0 | 0 |
| | 4 | 2 | 1 | 8 | 15 | 13 | 6 | 3 | 1 | 2 | 0 | 0 |
| | 5 | 1 | 2 | 2 | 2 | 19 | 13 | 9 | 8 | 6 | 1 | 1 |
| | 6 | 0 | 0 | 5 | 7 | 10 | 12 | 11 | 9 | 5 | 4 | 2 |
| | 7 | 0 | 0 | 3 | 3 | 4 | 10 | 9 | 9 | 12 | 7 | 1 |
| | 8 | 1 | 0 | 1 | 2 | 4 | 2 | 8 | 4 | 10 | 9 | 6 |
| | 9 | 0 | 0 | 0 | 1 | 2 | 0 | 5 | 5 | 6 | 7 | 4 |
| | 10 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 0 | 4 | 8 | 4 |

```
Chi-Quadrat-Test auf Homogenitaet der 2 Item-Gruppen
Chi-Quadrat= 101.6839  df=99  p= 0.4126  *)

*) ein p-Wert groesser 0.05 bedeutet, dass die beiden Item-Gruppen homogen sind
```

#### P14.8.7.8 Antwortmuster ausgeben

Eine Itembatterie bestehe aus 5 Items, z.B. 5 Mathe-Tests. Eine Person hat den 2. Test "voll richtig" gelöst und erhält dafür 2 Punkte. Die anderen Tests konnte die Person nicht lösen. Das Antwortmuster (engl. pattern) der Person ist dann

```
0 2 0 0 0
```

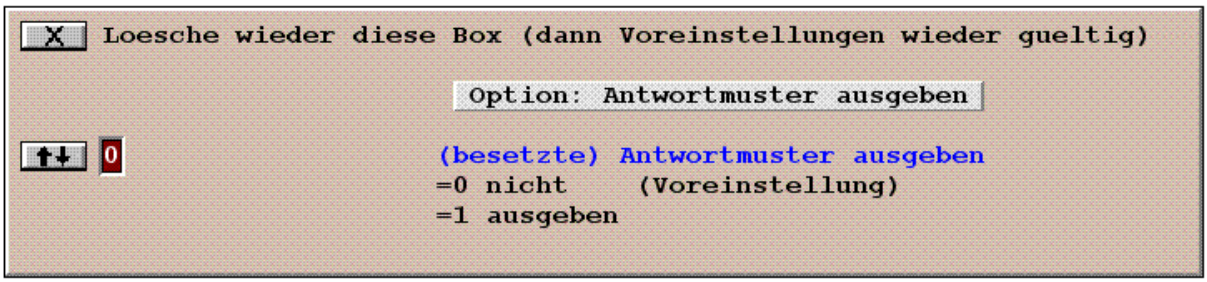
Die Zahl aller möglichen Antwortmuster ist

```
a1 * a2 * a3 * .... *am
```

```
m                =Zahl der Items
a1, a2, ... am =Zahl der Ausprägungen  des Item 1, 2, ... m
```

Im Almo-Rasch-Programm können nun die besetzten Antwortmuster ausgeben werden.

Die geöffnete Eingabebox ist folgende




Almo gibt dann die (mit Personen) besetzten Antwortmuster aus, z.B. so

**Antwortmuster**

```
 D  =Dezimalkodierung des pattern
 G  =Gesamtpunkte des pattern
 H  =Besetzungshaeufigkeit des pattern
 tH =theoretische Besetzungshaeufigkeit des pattern

  Nr.     D     G     H     tH     pattern
-----  ----  ----  ----  -------   ----------
    1     0     0     7    7.000   0 0 0 0 0 0
    2     1     1     4    4.015   0 0 0 0 0 1
    3     4     1    10    9.434   0 0 0 0 1 0
    4     5     2     4    7.172   0 0 0 0 1 1
    5     6     3     1    0.843   0 0 0 0 1 2
    .     .     .     .      .     . . . . . .
    .     .     .     .      .     . . . . . .
    .     .     .     .      .     . . . . . .
    .     .     .     .      .     . . . . . .
  217   570    10     2    5.242   1 1 2 2 2 2
  218   571    11     4    4.410   1 1 2 2 2 3
  219   573    10     2    2.630   1 1 2 2 3 1
  220   574    11     5    5.746   1 1 2 2 3 2
  221   575    12     3    3.000   1 1 2 2 3 3
```

Mit der Dezimalkodierung D kann für jedes pattern eine spezifische Dezimalzahl angegeben werden. Siehe dazu auch im Almo-Ordner "Algorithmen_in_C" das Unterprogramm "PatDez" im Modul "up_algorith11.c" sowie das Unterprogramm "DezPat" das die Dezimalzahl wieder zurück in ein pattern wandelt.

Die theoretische Besetzungshaeufigkeit ist jene, die gemäß dem Rasch-Modell zu erwarten wäre.

Almo gibt dann noch folgenden Hinweis aus.

```
******* von den 221 besetzten Antwortmuster sind
        163 mit weniger als 5 Personen besetzt
******* von den 221 besetzten Antwortmuster sind
        178 mit weniger als 5 Erwartungswerten besetzt
******* Wird der pauschale Modelltest gerechnet, dann
        entstehen valide Ergebnisse nur durch den Bootstrap-Test
```

Die Differenz zwischen beobachteter und theoretischer Besetzungshäufigkeit wird verwendet um den Pearson-Chi-Quadrat- und den Cressie-Read-Chi-Quadrat-Wert zu ermitteln. Diese beiden Chi-Quadrat-Werte bilden die „pauschalen“ Modelltests. Siehe dazu Abschnitt P14.8.8.5.

```
Pearson-Chi-Quadrat = Summe((H-tH)*(H-tH)/tH)
Cressie-Read = Summe(1.8*H* ((H/tH)**(2/3) -1.0))

jeweils summiert über alle besetzten pattern
** = potenzieren
```

### P14.8.7.9 Skalenwerte (Fähigkeitswerte) in Datei geben

Der Zweck einer Rasch-Analyse einer Itembatterie ist es letztendlich, für die Untersuchungspersonen Fähigkeitswerte zu ermitteln und jeder einzelnen Person ihren Fähigkeitswert zuzuweisen. Almo verfährt folgendermaßen: Es wird eine neue Datei angelegt. Die Originaldatei bleibt unberührt. In die neue Datei wird die Originaldatei übertragen und für jede Person hinter der seitherigen letzten Variablen der Gesampunkt und der Fähigkeitswert der Person angefügt.

Die geöffnete Eingabebox ist folgende

*1. Eingabefeld:*
Geben Sie einen neuen Dateinamen ein
Beispiel:

```
"C:\Almo15\PROGS\Skalwert"
```

Folgende 2 Bedingungen müssen erfüllt sein

1. Der Dateiname darf keiner Erweiterung wie z.B. ….fre besitzen, so wäre

```
"C:\Almo15\PROGS\Skalwert.fre"
```
falsch

Almo setzt selbst die Erweiterung .fre

2. Es muss der volle Pfad- und Dateiname angegeben werden.
   Sofern die Skalenwerte in einen Unterordner innerhalb des Almo-Ordners gespeichert werden sollen, dann ist folgende Abkürzung möglich

```
".\PROGS\Skalwert"
```

`C:\Almo15` wird durch einen Punkt ersetzt

Wenn Sie einen Dateinamen in das Eingabefeld schreiben, dann

1. erzeugt Almo zwei Dateien mit diesem Namen
   einmal im Format FREI und einmal im Format DIREKT
   Der Namen der beiden Dateien (im Beispiel) ist dann

   `".\PROGS\Skalwert.fre"` und
   `".\PROGS\Skalwert.dir"`

2. Und speichert in diese die Variablen aus der alten Datei,
   die in der Eingabebox "Datei aus der gelesen wird" angegeben ist

3. und speichert die im Rasch-Verfahren ermittelten Gesamtpunkte
   und die Fähigkeitswerte als die beiden letzten Variablen
   hinter die Variablen aus der alten Datei.

   Siehe nachfolgende Anmerkung zur Kodierung fehlender Wert
   Siehe nachfolgende Anmerkung zum Gesamtpunkt

4. weiterhin erzeugt Almo eine Datei der Variablennamen.
   Diese enthält
   - die Variablennamen aus der alten Datei (einschliesslich der eventuell in der Box "Freie Namensfelder" hinzugefügten oder modifizierten Namen
   - den Name "GesamtPunkt" und
   - den Name "Skala.." für die neue angehängte Fähigkeit- bzw. Skalenwertvariable, wobei Almo anstelle der 2 Punkte die Variablennummer der Skalenwertvariablen einsetzt.

Nachdem das Programm gelaufen ist, können Sie die Datei der Variablennamen in ein Fenster laden und Variablen- und Ausprägungsnamen beliebig verändern. Danach wieder speichern

Almo hat beispielsweise selbständig folgenden Namen geschrieben

```
Name 20=GesamtPunkt;
Name 21=Skala21;
```

Die Ziffer hinter "Skala.." (im Beispiel '21') ist die Variablennummer der Skalenwertvariablen.

Sie können diesen Variablenname beliebig verändern; z.B. so

```
Name20=GesPkt_MatheTest;
Name21=Faehigkeit_MatheTest
```

Der Name muss eindeutig sein, d.h. er darf kein 2. Mal auftreten.

*2. Eingabefeld:*

Die zu speichernden Fähigkeitswerte können umkodiert werden.
Beispiel: Mit

```
Runde 0.001
```

werden die Fähigkeitswerte auf die 3. Kommastelle gerundet. Aus beispielsweise `2.1309` wird `2.131`. Gerundet wird auf die Kommastelle an der in der „Runde"-Anweisung die 1 steht

Oder: `*100; Runde 1`

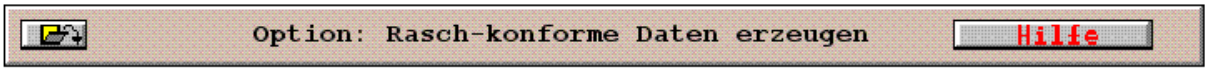
Zuerst wird mit 100 multipliziert, dann auf die Einer-Stelle gerundet. Aus beispielsweise 2.147 wird 215

*Anmerkung zur Kodierung fehlender Wert beim Speichern der neuen Datei*
Wie werden fehlende Werte in den beiden Dateien "xxxx.fre" und "xxxx.dir" die Almo anlegt geschrieben ? In der Datei im (lesbaren) freien Format xxxx.fre werden fehlende Werte durch die beiden Buchstaben kw oder KW kodiert, in der Datei im Format DIREKT xxxx.dir durch die riesige negative Zahl -1.0e36. Werden in nachfolgenden Analysen diese Dateien eingelesen, dann erkennt Almo diese Codes als "Kein-Wert-Code" und nimmt die von Almo vorgesehene oder vom Benutzer eingestellte Kein-Wert-Behandlung vor.

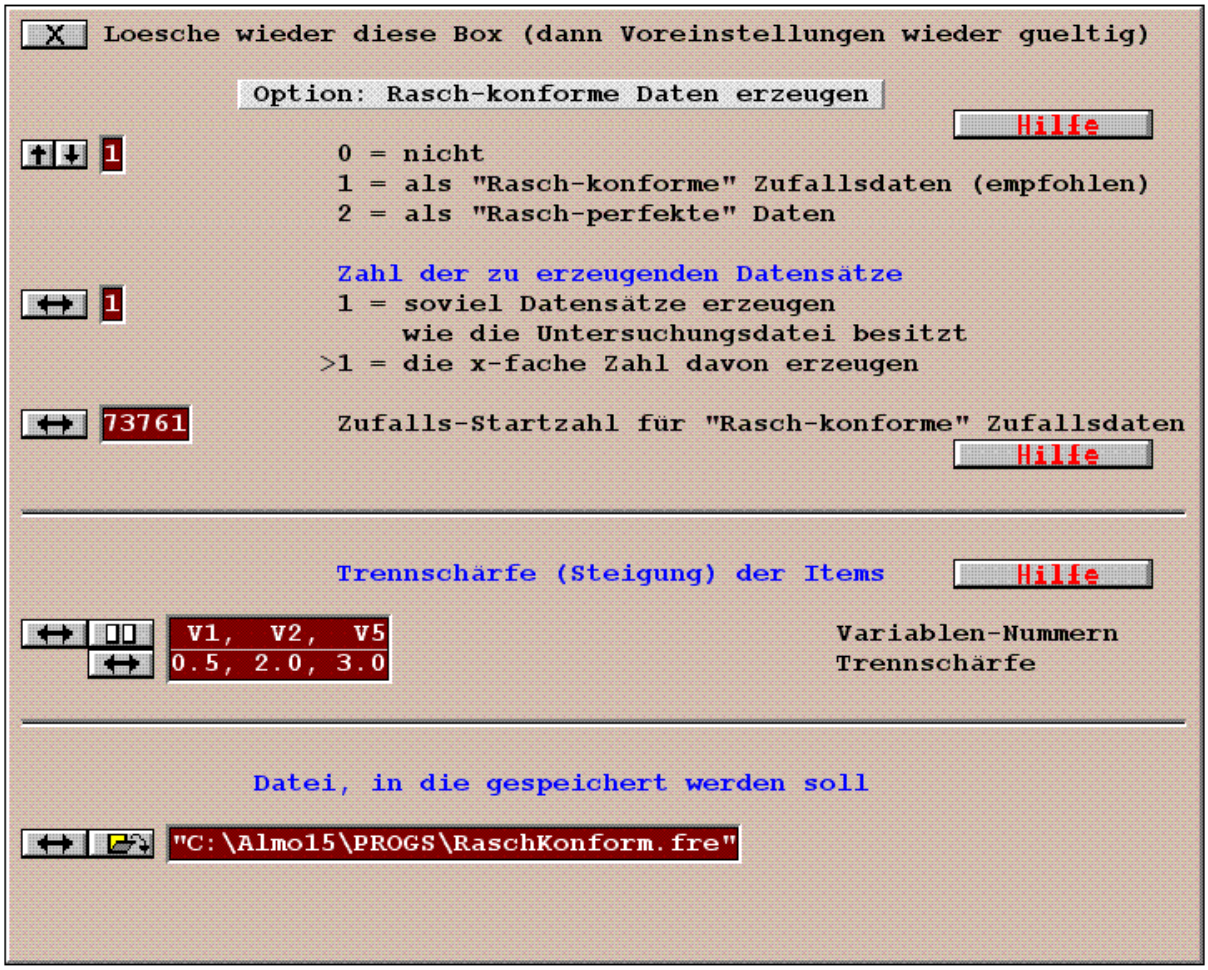
*Anmerkung zum Gesamtpunkt*
Der Gesamtpunkt, der in die neue Datei gespeichert wird ist jener der im Rasch-Verfahren errechnet wird Das geschieht in folgender Weise:

1. Die Items werden auf Werte-Untergrenze 0 transformiert.
   Das geschieht sehr einfach dadurch, dass von jedem Item-Wert
   der kleinste Wert des Items abgezogen wird.
   Der transformierte Wert wird für den Gesamtpunkt verwendet
2. Der Benutzer hat eventuell Items umkodiert. Er hat z.B.
   zwei benachbarte Ausprägungen zusammengefasst.
   Der umkodierte Wert wird für den Gesamtpunkt verwendet
3. Manchen Personen besitzen nicht in allen Items valide Werte
   Der Benutzer hat eine von Almo angebotene Methode gewählt,
   für einen fehlenden Wert einen Ersatzwert einzusetzen.
   Der Ersatzwert wird für den Gesamtpunkt verwendet
4. Erst nach diesen Bearbeitungen werden die Item-Werte
   zum Gesamtpunkt addiert.

### P14.8.7.10 „Rasch-konforme“ oder „Rasch-perfekte“ Daten erzeugen und in eine Datei geben




Geöffnete Optionsbox:




Almo verwendet die Rasch-Parameter der Untersuchungsdaten um eine Datei von künstlichen Personen zu erzeugen, die "Rasch-konform" bzw. "Rasch-perfekt" sind. Almo unterscheidet zwischen diesen beiden Formen. Diese Daten werden in eine Datei gespeichert. Für sie kann dann ein Rasch-Modell gerechnet werden. Dessen Ergebnisse können dann mit denen der Untersuchungsdaten verglichen werden. Besonders interessant ist dies für die Itemfit-Maße.

**Letztes Eingabefeld**
In dieses Eingabefeld muss der volle Pfad- und Dateiname eingetragen werden, z.B.

"C:\Almo\PROGS\RaschKonform.fre"

Wenn in einen Unterordner innerhalb des Almo-Ordners (z.B. "Progs") gespeichert werden sollen, dann ist eine Abkürzung möglich, z.B.

".\PROGS\RaschKonform.fre"

Anstelle von "C:\Almo" wird ein Punkt gesetzt

**1 und 2. Eingabefeld**

*"Rasch-konforme" Zufallsdaten erzeugen*

Wird in das 1. Eingabefeld eine 1 eingesetzt, dann erzeugt Almo "Rasch-konforme" Zufallsdaten nach folgender Vorgehensweise. Siehe dazu auch im Almo-Ordner "Algorithmen_in_C" das Unterprogramm "simord2" im Modul "up_algorith11.c" oder das Unterprogramm "SIMord" im Almo-Unterordnern "Rasch_in_R".

Aus den Untersuchungsdaten werden folgende Parameter übernommen:

Zahl der Items,
Zahl der Ausprägungen je Item,
Schwierigkeitsparameter,
Fähigkeitsparameter,
Personenzahl je Gesamtpunkt.

Die Schwierigkeits- und Fähigkeitsparameter werden in die logistische Grundgleichung (1) des Rasch-Modells eingesetzt. Durch gleichverteilte Zufallszahlen werden dann die Werte der Items je (künstliche) Person erzeugt. Siehe die Formeln in Abschnitt P14.93 und P14.8.9.8. Es wird also eine N*m-Datei erzeugt, die der empirischen Datei äquivalent ist – mit dem entscheidenden Unterschied, dass die simulierten Daten "Rasch-konform" sind.

Wird im **2. Eingabefeld** eine 1 eingesetzt, dann werden N Datensätze, also soviele Datensätze (Personen) erzeugt, wie die originale Untersuchungsdatei umfasst. Wird eine Zahl x grösser 1 eingesetzt, dann wird das x-fache an Datensätzen erzeugt. Wird z.B. 100 eingesetzt dann werden 100 mal soviele Datensätze erzeugt wie die originale Untersuchungsdatei umfasst. Die eingesetzte Zahl kann eine Kommazahl sein,

Natürlich gilt, dass die erzeugten Daten umso "Rasch-konformer" sind je mehr Datensätze erzeugt werden.

Werden für die 20 Pisa-Mathe-Items "Rasch-konforme" Zufallsdaten mit Multiplikator 100 erzeugt und diese dann dem Rasch-Verfahren unterworfen, dann entstehen Schwierigkeits-und Fähigkeits-Parameter, die mit denen aus der originalen Untersuchungsdatei bis zur 3. Kommastelle übereinstimmen. D.h. die für die Simulation vorgegebenen Schwierigkeits-und Fähigkeitsparameter werden reproduziert.

*"Rasch-perfekte" Daten erzeugen*

Wird in das 1. Eingabefeld eine 2 eingesetzt, dann erzeugt Almo "Rasch-perfekte" Daten nach

folgender Vorgehensweise. Aus den Untersuchungsdaten werden dieselben Parameter (wie oben beschrieben) übernommen. Siehe dazu auch im Almo-Ordner "Algorithmen_in_C" das Unterprogramm "simPerfekt_bed2" im Modul "up_algorith11.c" oder das Unterprogramm "SIMexakt" im Almo-Unterordnern "Rasch_in_R".

Es werden alle möglichen Antwortmuster gebildet. Eines nach dem anderen. Für jedes Antwortmuster wird entsprechend der grundlegenden logistischen Rasch-Formeln (1) aus P14.8.3 die Wahrscheinlichkeit seiner Besetzung mit Personen errechnet. Diese Wahrscheinlichkeit kann sehr klein werden. In die Rasch-Formeln werden die aus den Untersuchungsdaten ermittelten Schwierigkeits- und Fähigkeitsparameter eingesetzt.

Im **2. Eingabefeld** kann eine Kommazahl eingesetzt werden. Wird im 2. Eingabefeld eine 1 eingesetzt, dann wird diese Wahrscheinlichkeit mit der Häufigkeit der Untersuchungspersonen je dem Gesamtpunkt (der dem Antwortmuster entspricht) zu einer ganzzahligen Häufigkeit multipliziert. Durch die Rundung auf eine Ganzzahl kann eine 0 entstehen - also Information verloren gehen. Wird aber etwa 100 eingesetzt, dann werden die empirische Häufigkeiten "verhundertfacht". Dann gehen weniger Informationen verloren. Es werden aber immer noch 0-Häufigkeiten entstehen.

Ein Beispiel:

```
Das Antwortmuster einer 5-Item-Skala sei           1 0 1 0 0.
     dementsprechend ist der Gesamtpunkt           2
die empirische Häufigkeit bei 2 Gesamtpunkten sei  5
als Wahrscheinlichkeit der Besetzung des
           Antwortmusters sei entstanden         p=0.012.
```

Die Besetzungshäufigkeit für die "Rasch-perfekten" Daten ist dann

```
0.012 * 5 = 0.06    auf Ganzzahl gerundet 0
```

Es werden Daten verloren. Es kann keine wirklich "Rasch-perfekte" Datei entstehen. Wird im 2. Eingabefeld als Multiplikator 100 eingegeben dann entsteht

```
0.012 * 5 * 100 = 6
```

Keine Daten gehen verloren. Allerdings entsteht nun eine Datei mit der 100-fachen Größe der Untersuchungsdatei. Das kann unzumutbar sein. Wäre jedoch der p-Wert 0.0012, dann würde auch bei Multiplikator 100 eine 0 entstehen. Ein solch kleiner p-Wert kann entstehen, wenn sehr viele Antwortmuster möglich sind. Bei den 20 Pisa-Mathe-Items sind es z.B. 1048121.

*"Rasch-perfekte" Daten prüfen.*
Ob eine "Rasch-perfekte" Datei entstanden ist, kann überprüft werden. Wird mit den so entstandenen Daten ein Rasch-Modell gerechnet, dann müssen dieselben Schwierigkeits-und Fähidkeitsparameter (etwa bis zur 2 oder 3. Kommastelle) entstehen wie bei den originalen Untersuchungsdaten. Beim Andersen-Test (in grafischer Darstellung) muss die Diagonale eine Steigung von sehr nahe 1.0 besitzen und die Distanzen der Itempunkte von der Diagonalen müssen sehr klein sein.

Folgerung:

1. Je größer der in Eingabefeld 2 eingesetzte Multiplikator umso "perfekter" (!) werden die "Rasch-perfekten" Daten. Allerdings werden dabei Personenzahlen generiert, die um den Multiplikator vervielfacht sind.

2. Je mehr Items eine Itembatterie umfasst und je mehr Kategorien die Items haben - umso mehr mögliche Antwortmuster entstehen dann und umso kleiner werden die Besetzungs-Wahrscheinlichkeiten mit der Folge, dass immer mehr Daten verloren gehen.

3. Bei kleiner Item- und Kategorienzahl und nicht zu vielen Personen kann das Verfahren funktionieren. Siehe dazu das Beispiel mit den 5 dichotomen KFT-Items (im Prog14m4). Wird der Multiplikator im 2. Eingabefeld auf 100 gesetzt, dann werden 30000 Datensätze generiert (in wenigen Sekunden). Werden diese dem Rasch-Verfahren unterworfen, dann entstehen Schwierigkeitsparameter, die bis zur 3. Kommastelle identisch mit denen der originalen 300 KFT-Daten sind und Fähigkeitswerte, die bis zur 4. Kommastelle identisch mit den originalen sind.

   Werden hingegen für die 20 dichotomen Mathe-Items aus Pisa 2009 für Austria "perfekte" Daten mit Multiplikator 100 erzeugt, dann differieren die Schwierigkeits- und Fähigkeitsparameter der simulierten und der originalen Daten schon an der 1. Kommastelle mit bis zu 0.5. Die erzeugten Daten sind keine "Rasch-perfekten" Daten.

Zusammengefasst: Wenn die Zahl aller möglichen Antwortmuster sehr groß ist, dann versagen die "Rasch-perfekten" Daten während die "Rasch-konforme" Zufallsdaten ein genügend gutes Ergebnis erbringen. Bei kleiner Zahl möglicher Antwortmuster erbringen die "Rasch-perfekten" Daten ein geringfügig besseres Ergebnis.

*Welche simulierten Daten sind besser ?*
Allgemein gilt, dass der Benutzer kein Risiko eingeht, wenn er sich für die "Rasch-konformen" Zufallsdaten entscheidet.

**4. Eingabefeld: Trennschärfe (Steigung) der Items**

Bleibt das Eingabefeld leer, dann setzt Almo die Trennschärfe (Steigungen) aller Items auf 1.0. Dies sollte der Normalfall sein.

Dieses Eingabefeld dient ausschliesslich dem Experimentieren. Das Rasch-Modell fordert, dass alle Items dieselbe Trennschärfe besitzen. In Almo wird dies hauptsächlich durch Outfit und Infit überprüft

Wird der "Trennschärfe-Parameter" explizit in die die Formel für die dichtomen Items eingeschrieben, dann müsste also korrekterweise die Formel so geschrieben werden:

$$p = \frac{1}{1+e^{(c*(s-f))}}$$

```
e  = e-Zahl 2.718...
p  = Wahrscheinlichkeit einer positiven Antwort
s  = Schwierigkeitsparameter
f  = Fähigkeitsparameter
c  = Steigungsparameter der logistischen Kurve
ap = Zahl der Item-Kategorien minus 1
```

c ist der Steigungsparameter der logistischen Kurve. c kann als "Trennschärfeparameter" bezeichnet werden. Die Steigung der Tangente im Wendepunkt der Kurve ist c/4.

Im Rasch-Modell muss c für alle Items denselben Wert besitzen. Anders im "Birnbaum-Modell", das die obige erweiterte Gleichung verwendet und es explizit erlaubt, dass die Steigungen der logistischen Funktion der Items verschieden sein dürfen. Die Folge davon ist, dass sich die Kurven der Items überschneiden können.

Auch für polytome Items sollte c in die Formel aufgenommen werden. Dort müsste die Formel dann lauten:

$$p = \frac{e^{(c*(j*f-s))}}{sum}$$ für Itemkategorie j

wobbei

$$sum = \sum_{k=0}^{ap} e^{(c*(k*f-s))}$$ Summierung von k=0 bis k=ap

Für alle Items muss c im Rasch-Modell denselben Wert besitzen. Üblicherweise wird deswegen in dieser Formel c weggelassen. D.h. dass c auf 1 gesetzt wird.

Dem Almo-Benutzer wird nun die Möglichkeit geboten, Daten zu erzeugen mit Items, die unterschiedliche Steigungen besitzen. Er kann dann kontrollieren, wie sich z.B. die Itemfit-Koeffizienten verhalten.

Werden, wie in nachfolgender Eingabebox, ungleiche Steigungen eingegeben, dann entstehen Daten, die selbstverständlich nicht mehr raschkonform sind. Wird danach mit diesen Daten (in unserem Beispiel heißen sie ".\Progs\Raschkonform.fre" ) ein Rasch-Analyse gerechnet, dann unterstellt das Modell und entsprechend auch unser Programm für alle Items gleiche Steigungen mit c=1 und berechnet so die Schwierigkeits- und Fähigkeitswerte und die sonstigen Koeffizienten. Auch in den Grafiken, die unser Programm ausgibt, werden die logistischen Kurven mit Steigung c=1 gezeichnet. Die In- und Outfit-Koeffizienten, die das Programm ausgibt, weisen jedoch nach, dass nicht-raschkonforme Items analysiert wurden.

Die Eingabe soll an nachfolgendem Beispiel erläutert werden

a. Wenn beide Eingabefelder leer sind, dann setzt Almo den Steigungs-Parameter c für alle Items auf 1.0.
b. Im abgebildeten Beispiel werden im oberen Eingabefeld die 3 Items mit den Variablen-Nummern 1, 2 und 5 eingetragen. Insgesamt, so wollen wir annehmen seien die Items V1 bis V10 dem Rasch-Modell unterworfen worden.
c. Im unteren Eingabefeld werden nun für die 3 selektierten Items folgende Steigungen c eingetragen: Für

V1 0.5
V2 1.0
V5 3.0

Die nicht-selektierten Items V3,4,6 bis 10 werden von Almo automatisch auf 1.0 gesetzt.
d. Im oberen Eingabefeld werden die Variablen-Nummern (mit oder ohne vorausgehendes V) geschrieben. Zulässig sind anstelle der Nummern auch Variablen-Namen. Im unteren Eingabefeld werden, korresponierend zu den Variablen-Nummern, die Steigungswerte c geschrieben.
e. Möglich sind auch folgende kurze Eingaben im oberen Feld, beispielsweise

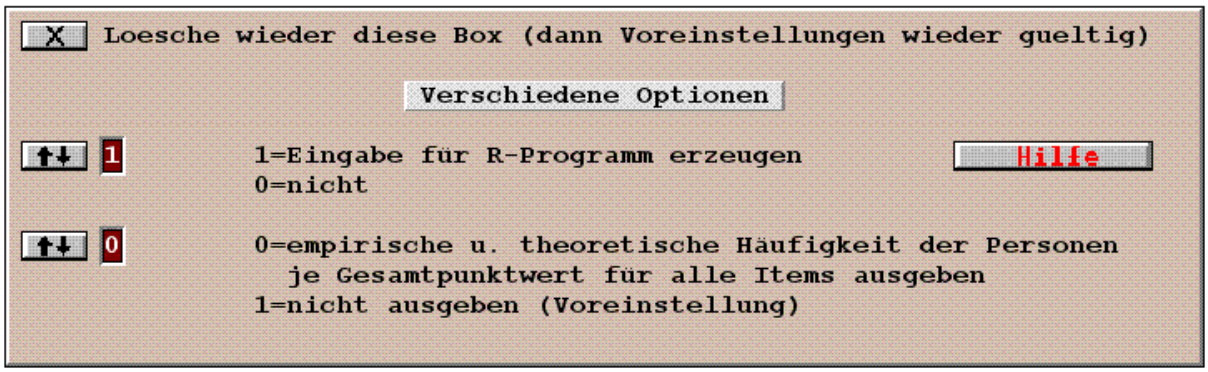
**V6:14** das sind die Variablen V6 bis V14
oder
**V3,7:11,13** das sind die Variablen V3,7 bis 11 und V13

Der Doppelpunkt bedeutet "fortlaufend bis...". Im unteren Eingabefeld müssen dann entsprechend gleich viele Steigunswerte c in korrespondierender Folge eingetragen werden
f. Wenn viele Items dieselbe Steigung erhalten sollen, dann kann so geschrieben werden

Im oberen Feld werden die Variablen V1 bis V20 angegeben. Im unteren Feld wird angegeben, dass 20 mal der Wert 1.5 verwendet werden soll. Der Stern * ist nicht alsMultiplikations-zeichen zu interpretieren. Er bedeutet, dass der Wert 1.5 20 mal wiederholt wird.

### P14.8.7.11 R-Programm erzeugen / Ausgabe erweitern

Wird die Optionsbox geöffnet, dann sieht man folgendes.




Im Almo-Ordner „Rasch_in_R“ sind die von Heinrich Potuschak geschriebenen R-Funktionen für das ordinale Rasch-Modell enthalten. Mit diesen Funktionen kann auch konkret gerechnet werden, sofern die Antwortmuster (pattern) der Personen in der Dezimalkodierung vorliegen und die pattern-Häufigkeiten bekannt sind. Wird in das Eingabefeld eine 1 eingesetzt, dann wird in die Ergebnisdatei von Almo ein Textabschnitt eingefügt, der für das R-Programm die Funktionsaufrufe und die dazu notwenigen Daten enthält.

Wenn der Benutzer dieselben Untersuchungsdaten (die er mit Almo gerechnet hat) nun auch mit R rechnen will, dann muss er nur folgendes tun.

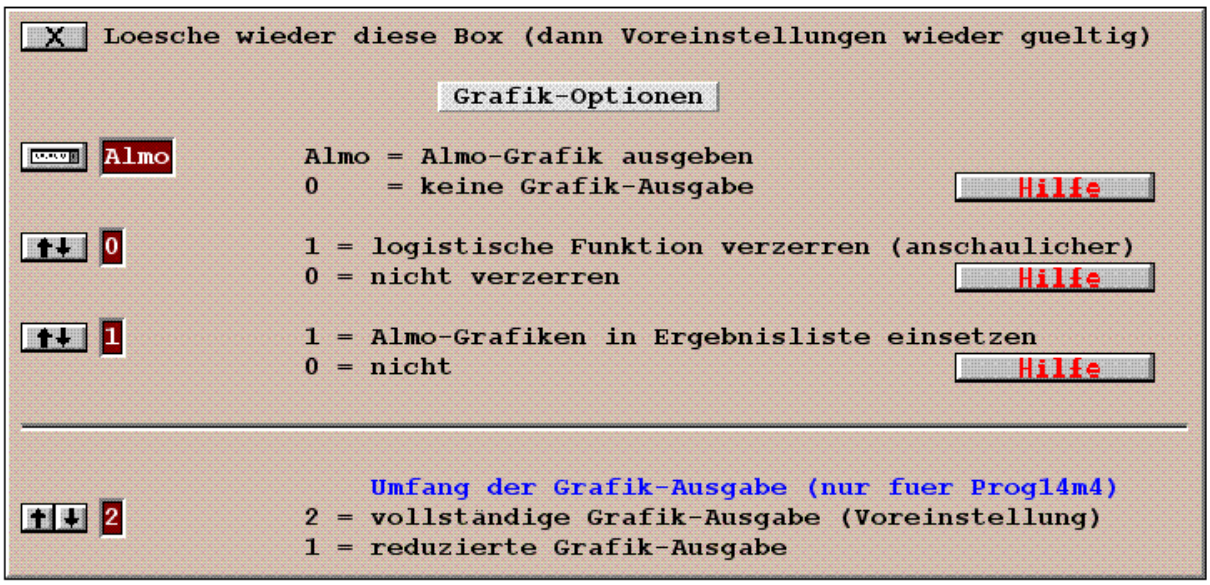
1. Almo starten. In Almo den Ordner „Rasch_in_R“ öffnen und die darin befindliche Datei **oRM.R** öffnen, d.h. in ein Fenster laden. oRM.R enthält alle R-Funktionen für das Rasch-Modell.
2. Den gesamten Text als Block markieren über das Menü „Bearbeiten / Markiere Alles“ und danach kopieren, wieder über das Menü „Bearbeiten / Markierten Text-Block in Zwischen-ablage“
3. R starten. In R ein „neues Skript“ erzeugen. Mit der Tastenkombination „Strg/V“ oRM.R in dieses Skript kopieren. Klick auf das Menü „Bearbeiten / Alles ausführen“. Damit stehen alle R-Funktionen zur Verfügung
4. In Almo das Rasch-Programm Prog14m4 mit den ausgewählten Daten und der geöffneten Optionsbox (wie oben) rechnen. In der Ergebnisliste ist dann ein Textabschnitt eingefügt, der die R-Funktionsaufrufe und die dazu notwenigen Daten enthält.
5. Diesen Textabschnitt in der Almo-üblichen Weise markieren, etwa auf die 1. Zeile mit gedrückter Strg-Taste klicken, dann auf die letzte.
6. Mit der Tastenkombination "Strg", dann "C" wird der Textblock kopiert (= in die Windows Zwischenablage geschoben)
7. In R wird dann ein zweites „neues Skript“ erzeugt und mit "Strg" und "V" der Textblock aus der Zwischenablage wieder geholt und in das erzeugte Fenster gestellt. Dann kann mit Klick auf „Bearbeiten / Alles ausführen“ gerechnet werden.

**Empirische und theoretische Häufigkeiten ausgeben**

In obiger Optionsbox kann (im 2.Eingabefeld) die Ergebnisliste erweitert werden durch umfangreiche Tabellen, die die empirischen und (durch das Rasch-Modell prognostizierten) theoretischen Häufigkeiten je Gesamtpunkt für die einzelnen Items enthalten. Dieser Ausgabeteil kann seh umfangreich werden.

### P14.8.7.12 Grafik-Optionen

Wird die Box geöffnet sieht man folgendes




***Eingabefeld 1:***
1=Almo-Grafik erzeugen
0=nicht

***Eingabefeld 2:***
1= logistische Kurven und Kategorienkurven verzerrt darstellen (voreingestellt)
0= nicht; darstellen mit Trennschärfeparameter =1
Almo verzerrt standardmäßig die Grafiken der logistischen Kurven der dichotomen Items bzw. der Kategorienkurven der polytomen Items. Der Hintergrund für diese Vorgehensweise ist folgender:

Die Formel der logistischen Funktion für dichotome Items lautet

$$p = \frac{1}{1+e^{(c*(s-f))}}$$

```
e = e-Zahl 2.718...
p = Wahrscheinlichkeit einer positiven Antwort
s = Schwierigkeitsparameter
f = Fähigkeitsparameter
c = Steigungsparameter der logistischen Kurve =1.0
```

Der Steigungsparameter c ist im Rasch-Modell eine Konstante für alle Items. c wird üblicher-weise in die Formel nicht aufgenommen bzw. =1.0 gesetzt. Damit wird die Bedingung der "gleichen Trennschärfe" aller Items realisiert. Siehe die ausführliche Darstellung im nach-folgenden Abschnitt P14.8.8.4.2.1 Die Steigung der logistischen Rasch-Funktion

Die Steigung der Tangente im Wendepunkt der logistischen Kurve ist

$$c = 1$$

Das sind 45 Grad. Die logistische Kurve des Rasch-Modells verläuft relativ flach, so dass der eigentlich S-förmige verlauf nicht richtig sichtbar wird.

Das gilt auch mit entsprechenden Abwandlungen für die Kategorienkurven der polytomen Items.

Um die Grafik anschaulicher zu machen verzerrt Almo standardmäßig die Grafik, indem die vertikale Koordinatenachse verlängert wird. Allerdings beträgt die Steigung der Kurve dann deutlich mehr als 45 Grad. Dafür wird der S-förmige Verlauf der logistischen Kurve deutlicher sichtbar.

Der Benutzer kann jedoch, wenn er im Eingabefeld die 0 einsetzt, diese Verzerrung verhindern.

***Eingabefeld 3:***
Wenn Sie '1' eingeben, dann werden die Almo-Grafiken direkt in die Ergebnisliste eingesetzt. Wenn Sie durch die Ergebnisliste blättern oder scrollen, dann werden Ihnen (anschliessend an Tabellen und Matrizen) auch die Almo-Grafiken gezeigt.

Wenn Sie '0' eingeben, dann können die Almo-Grafiken nur im Grafik-Editor angeschaut werden. In der Ergebnisliste ist dann (anschliessend an Tabellen und Matrizen) nur ein Grafikknopf enthalten. Durch Klick auf diesen Knopf gelangen Sie in den Grafik-Editor, wo Ihnen die Grafik gezeigt wird..

***Eingabefeld 4:***
Umfang der Grafik-Ausgabe
2 = vollständige Grafik-Ausgabe (Voreinstellung)
1 = reduzierte Grafik-Ausgabe

Bei 1 werden die Schwellenfunktionen für jedes Item gezeigt Bei 2 werden zusätzlich die Gesamtpunkte je Ausprägung in die Grafiken eingezeichnet. Dadurch entstehen u.U. sehr viele Grafiken. Der Vorteil ist jedoch, dass der Benutzer einen "optischen" Eindruck erhält, wie gut sich die Gesamtpunkte an die logistischen Kurven anschmiegen. Siehe dazu die Grafik in obigem Abschnitt P14.8.3 „Die Kategorienfunktion im ordinalen Rasch-Modell“

Gezeigt werden die 3 Kurven eines 3-kategoriellen Items. Dabei sind die Gesamtpunkte für die Kategoriekurve 1 eingezeichnet. Sie schmiegen sich sehr eng an die theoretische Rasch-Kurve an.

### P14.8.8 Ausgabe aus Prog14m4

Eine ausführliche und sehr übersichtliche Erläuterung der verschiedenen Koeffizienten, die Prog14m3 und Prog14m4 ausgeben, findet man im Lehrbuch von Rost..
Jürgen Rost:Lehrbuch Testtheorie - Testkonstruktion, 2004, insbesondere S. 115 ff
und S.203 ff

Die im folgenden dargestellte Ergebnis-Ausgabe entstand aus Prog14m4 mit den simulierten Daten „simRasch815.fre“. Das Programm wird erreicht durch Klick auf den Knopf „Verfahren“, dann Eintrag „Rasch-Modell“ und Prog14m4.

### P14.8.8.1 Ausgabe aus Almo: Basis-Statistiken

```
Zahl der eingelesenen Datensaetze:      815
                                                                                Kein-Wert
                                                                                Faelle
Variable            mittlerer Wert    Streuung    Untergrenze     Obergrenze     abs   in %
--------            --------------    --------    -----------     ----------    -----------
  1 Item_A                 -             -           0.0000         1.0000      0    0.00
  2 Item_B                 -             -           0.0000         1.0000      0    0.00
  3 Item_C                 -             -           0.0000         2.0000      0    0.00
  4 Item_D                 -             -           0.0000         2.0000      0    0.00
  5 Item_E                 -             -           0.0000         3.0000      0    0.00
  6 Item_F                 -             -           0.0000         3.0000      0    0.00
```

Almo transformiert für das Rasch-Verfahren automatisch alle Items auf Untergrenz=0. Hier wird aber noch einmal mitgeteilt, was die Werte-Unter- und Obergrenzen tatsächlich sind und wie viele Kein-Wert-Fälle aufgetreten sind.

### P14.8.8.2 Schwierigkeits-, Schwellen- und Fähigkeitsparameter

```
Ergebnisse aus Almo: Ordinales Rasch-Modell
-------------------------------------------
Verwendete Schätzmethode: Bedingte Maximum-Likelihood-Schätzung
Vorgehensweise wenn Werte fehlen: Datensatz insgesamt ausschliessen

Zahl der eingelesenen Datensaetze  = 815
Zahl der verarbeiteten Datensaetze = 815

Zahl der Personen je Gesamtpunktwert
Gesamtpunktwert     Zahl der Personen
      0                    7
      1                   27
      2                   55
      3                   84
      4                  124
      5                  113
      6                  130
      7                  111
      8                   58
      9                   54
     10                   31
     11                   18
     12                    3

Mittelwert=   5.5227
Median    =   5.0000

                              Personenzahl je Auspraegung
Variable                    absolut                  relativ
   V1  Item_A
      0                       171                    0.209816
      1                       644                    0.790184
   V2  Item_B
```

```
      0                        293                    0.359509
      1                        522                    0.640491
   V3  Item_C
      0 schwach                249                    0.305521
      1 mittel                 410                    0.503067
      2 stark                  156                    0.191411
   V4  Item_D
      0 wenig                  281                    0.344785
      1 etwas                  401                    0.492025
      2 viel                   133                    0.16319
   V5  Item_E
      0 nie                    107                    0.131288
      1 selten                 472                    0.579141
      2 manchmal               164                    0.201227
      3 oft                     72                    0.0883436
   V6  Item_F
      0 sehr leicht            187                    0.229448
      1 eher leicht            388                    0.476074
      2 eher schwer            178                    0.218405
      3 sehr schwer             62                    0.0760736


Bedingtes Log-Likelihood:  -2388.613269
nach 12 Iterationen
und letzter Iterationsdifferenz von 0.00005117



Schwierigkeitsparameter
                            0            1            2            3
                        ---------    ---------    ---------    ---------
   V1  Item_A               -        -1.780127
   V2  Item_B               -        -0.888140
   V3  Item_C               -        -0.981801     0.281247
   V4  Item_D               -        -0.785696     0.668386
   V5  Item_E               -        -2.219833    -1.146065     0.405705
   V6  Item_F               -        -1.365533    -0.503523     1.312928



Schwellenparameter
                            0            1            2            3
                        ---------    ---------    ---------    ---------
   V1  Item_A               -        -1.780127
   V2  Item_B               -        -0.888140
   V3  Item_C               -        -0.981801     1.263048
   V4  Item_D               -        -0.785696     1.454082
   V5  Item_E               -        -2.219833     1.073768     1.551770
   V6  Item_F               -        -1.365533     0.862010     1.816451

Summe der Schwellenparamter =    0.0000  (nullsummennormiert)
```

************Erläuterung**

Für dichotome Variable sind Schwierigkeits- und Schwellenparameter gleich. Bei polytomen Items sind die Schwierigkeitsparameter ein Zwischenergebnis, das selbst nicht interpretiert wird. Bedeutsam sind die "Schwellenparameter". Sie werden im nachfolgenden Abschnitt in Zusammenhang mit der Grafik der "Kategorienfunktion" ausführlich erläutert Der Begriff des "Schwellenparameters" soll anhand der Grafik interpretiert werden, die Almo weiter unten in der Ausgabeliste ausgibt. Dabei wird die Grafik für V3 verwendet

**"Verdrehte" Schwellenwerte**

Bei der Analyse von 5 4-kategoriellen Items sind folgende Schwellenwerte ausgegeben worden:

```
Schwellenparameter
                            0           1           2           3
                        ---------   ---------   ---------   ---------
   V1                       -       -0.084554    0.050098   -0.024163
   V2                       -       -0.049294   -0.043809    0.041587
   V3                       -        0.046343    0.054822   -0.018073
   V4                       -        0.018402    0.014446    0.002629
   V5                       -        0.001785   -0.005492   -0.004728
```

Almo bringt folgende Fehlermeldung, rechnet aber weiter:

```
                                          ******************** FEHLER
         Die Kategorien der Variable V1 sind nicht aufsteigend geordnet
         Die Kategorien der Variable V3 sind nicht aufsteigend geordnet
         Die Kategorien der Variable V4 sind nicht aufsteigend geordnet
         Die Kategorien der Variable V4 sind nicht aufsteigend geordnet
         Die Kategorien der Variable V5 sind nicht aufsteigend geordnet
```

Die Schwellenwerte müssen ansteigen. Bei V1 liegt die 3. Schwelle unterhalb der 2. Schwelle. Es gibt 3 Möglichkeiten zu reagieren

1. Die Variable wird aus der Analyse ausgeschlossen

2. Die Kodierung der Ausprägungen wird durch eine Umkodierung entsprechend "umgedreht", etwa so

```
V1 ( 2=3; 3=2 )
```

Das muss natürlich inhaltlich sinnvoll sein. Diese Anweisung wird in die Umkodierungsbox geschrieben. Siehe oben Abschnitt **P14.8.7.3.**

3. Die Kategorien, die "verdrehte" Schwellenwerte besitzen, werden zusammengefasst. In unserem Beispiel: Die Variable V1 besitzt die 4 Ausprägungen 0, 1, 2, 3. Die Kategorien 2 und 3 sind "verdreht" In der Umkodierunsbox wird folgende Umkodierung geschrieben

```
V1 ( 0=0; 1=1; 2,3 =2 )
```

Die erste beiden Anweisung 0=0 und 1=1 sind redundant, können also unterbleiben. Mit der 3. Anweisung 2,3=2 werden die Kategorien 2 und 3 zu 2 zusammengefasst.

Bei V4 im Beispiel liegt die Vermutung nahe, dass sie insgesamt verkehrt herum kodiert ist. Man kann versuchen durch die Umkodierung

```
V4 ( 0:3=Umdrehen )
```

die Variable insgesamt umzudrehen.

**Standardfehler der Schwierigkeitsparameter**

```
Standardfehler der Schwierigkeitsparameter
                            0           1           2           3
                        ---------   ---------   ---------   ---------
```

```
   V1  Item_A                -         0.094163
   V2  Item_B                -         0.082695
   V3  Item_C                -         0.072776     0.098975
   V4  Item_D                -         0.073106     0.104014
   V5  Item_E                -         0.075287     0.092092     0.132357
   V6  Item_F                -         0.073769     0.090572     0.138727
```

**Fähigkeitswerte**

```
Faehigkeitswerte
WLE (=weigthed likelihood estimates) nach T.A. Warm
Gesamtpunktwert    Faehigkeitswert    Standardfehler
      0               -4.050137          1.608362
      1               -2.740193          1.000489
      2               -1.999759          0.837238
      3               -1.411733          0.765966
      4               -0.881409          0.730809
      5               -0.372290          0.710488
      6                0.120199          0.693770
      7                0.582641          0.680562
      8                1.015815          0.679719
      9                1.444889          0.702227
     10                1.916143          0.765512
     11                2.527603          0.920989
     12                3.697937          1.503578

Mittelwert der Faehigkeitswerte =  -0.011561
Andrich's Reliabilitaet         =   0.635225
Korrelation Gesamtpunkt mit Faehigkeitswert = 0.994013
```

************Erläuterung**

Das Rasch-Modell findet für Personen, die aus der Beantwortung der 6 Items 0 Punkte erzielten, einen Fähigkeitswert von -4.050137; für Personen die 1 Gesamtpunkt erzielten eine Fähigkeitswert von -2.740193 usw. Die Fähigkeitswerte besitzen einen Standardfehler. Wir dürfen z.B. vermuten, dass der "wahre" Fähigkeitswert bei 1 Gesamtpunkt im Intervall von -2.740193 plus/minus 2*1.000489 liegt; dies mit 95%iger Sicherheit. Das gilt natürlich nur, wenn die Daten als Zufallsstichprobe gewonnen wurden. Der Standardfehler nimmt von den mittleren Gesamtpunkten in Richtung zu den Extremwerten 0 und 12 zu.

Der Fähigkeitswert ist vermutlich das wichtigste Ergebnis unserer Analyse. Der zentrale Zweck des Rasch-Verfahrens ist es, (für eine den Forscher interessierende Messdimension) für jede untersuchte Person, einen Fähigkeitswert zu finden, der in einem späteren Schritt an den Datensatz der Person als zusätzliche Variable angehängt werden kann. Siehe dazu Abschnitt **P14.8.7.9 .** Damit sind dann Analysen möglich, in die die Messdimension (durch den Fähigkeitswert) unmittelbar eingeht und nicht die einzelnen Indikatoren (die Items) der Messdimension. So ist es z.B. in der Pisa-Studie möglich den Zusammenhang von Mathematik-Begabung (ausgedrückt durch den Fähigkeitswert) und Geschlecht und soziale Herkunft usw. zu ermitteln - was kaum möglich ist, wenn die Mathematik-Begabung über einen der 22 Mathe-Items ausgedrückt wird.

**Fähigkeitswert oder Gesamtpunkt ?**

Verblüffend ist die extrem hohe Korrelation zwischen dem Gesamtpunktwert und dem Fähigkeitswert. In unserem Beispiel wurden "Rach-konform" simulierte Daten verwendet, so dass sich diese extreme Korrelation zum Teil daraus erklärt. Aber auch bei empirischen Daten - sofern sie annehmbar Rasch-konform sind - treten kaum Korrelationen unter 0.9 auf. Bei den Pisa-Mathe-

Daten in unserem Beispiel-Programm "Rasch_PisaMathe.Alm" beträgt die Korrelation 0.988557 und bei den 5 "Neoffi"-Daten, die bei Rost (2004) als Beispiel verwendet werden, 0.984928 (siehe unser Beispiel-Programm "Rasch_Neoffi.Alm"). Die beiden Beispielprogramme können aus der Programm-Maske "Prog14m4.Msk" heraus aufgerufen werden.

Es liegt somit der Gedanke nahe: "Wozu Rasch, wenn ich mit dem Gesamtpunkt dasselbe erreiche?“. Folgendes ist jedoch zu bedenken. "Rasch-konforme“ Items haben u.a. folgende Eigenschaften:

1. Sie unterscheiden sich in identifizierbarer Weise in ihrer Schwierigkeits- bzw. in ihren Schwellenwerten.

2. Sie besitzen annähernd gleiche Trennschärfe. Bei dichotomen Items bedeutet dies, dass sich die logistischen Kurven der Items **nicht** schneiden. Es kann also nicht der Fall eintreten, dass ein leichtes Item zwar von Personen geringer Fähigkeit gelöst wird, aber nicht signifikant oft, dass es von Personen hoher Fähigkeit nicht gelöst wird. Über "Outfit“ und "Infit“ wird auf Trennschärfe getestet. Diese Itemfitmaße leiden allerdings unter gewissen Problemen, auf die noch eingegangen wird.

3. Rasch-konforme Items messen auf einer gemeinsamen Dimension. Getestet wird das implizit über die Itemfit-Maße und über Personen- und Item-Homogenitätsmaße. Siehe nachfolgend.

4. Bei polytomen Items besteht die (relative) Sicherheit, dass ihre 0-1-2-3... kodierten Kategorien tatsächlich aufsteigende Schwellenwerte besitzen, so dass eine Gesampunktbildung erst möglich wird. Bei Simulationen mit "reinen", nicht auf Rasch-Konformität gewandelten Zufallsdaten treten regelmäßig "verdrehte" Schwellenwerte auf. Siehe unser Beispielprogramm "Rasch_Zufalldat2.Alm" (erreichbar über Menü "Almo/Beispielprogramm/Prog14".

Ein derart gebildeter Gesamtpunkt besitzt eine andere Qualität als ein Gesamtpunkt, der ohne Überprüfung gebildet wird. Im besonderen Fall, dass die Items unterschiedlich viele Kategorien besitzen ist eine Gesamtpunktbildung ohne Rasch-Modell gar nicht möglich.

Die Schlussfolgerung wäre also: Der Gesamtpunkt kann anstelle des Fähigkeitswerts problemlos verwendet werden - nachdem zuvor eine Überprüfung durch das Rasch-Modell stattgefunden hat. So kann das Rasch-Modell auch betrachtet werden als eine sorgfältige Untersuchung von Gesamtpunkten.

**Häufigkeitstabelle**

```
Haeufigkeiten der Auspraegungen je Gesamtpunkt
 G = Gesamtpunkt
 Z = Zahl der Personen je Gesamtpunkt
 F = Faehigkeitswert je Gesamtpunkt

  G     Z     F    |   V1    |   V2    |   V3         |   V4
```

| | | | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 1 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7 | -4.050 | 7 | 0 | 7 | 0 | 7 | 0 | 0 | 7 | 0 | 0 |
| 1 | 27 | -2.740 | 22 | 5 | 26 | 1 | 24 | 3 | 0 | 23 | 4 | 0 |
| 2 | 55 | -2.000 | 29 | 26 | 45 | 10 | 38 | 16 | 1 | 42 | 13 | 0 |
| 3 | 84 | -1.412 | 36 | 48 | 58 | 26 | 51 | 29 | 4 | 55 | 28 | 1 |
| 4 | 124 | -0.881 | 34 | 90 | 66 | 58 | 58 | 59 | 7 | 57 | 64 | 3 |
| 5 | 113 | -0.372 | 20 | 93 | 34 | 79 | 37 | 69 | 7 | 41 | 65 | 7 |
| 6 | 130 | 0.120 | 10 | 120 | 29 | 101 | 17 | 94 | 19 | 39 | 78 | 13 |
| 7 | 111 | 0.583 | 7 | 104 | 14 | 97 | 10 | 69 | 32 | 13 | 68 | 30 |
| 8 | 58 | 1.016 | 3 | 55 | 6 | 52 | 5 | 31 | 22 | 2 | 32 | 24 |
| 9 | 54 | 1.445 | 3 | 51 | 6 | 48 | 2 | 22 | 30 | 1 | 30 | 23 |
| 10 | 31 | 1.916 | 0 | 31 | 2 | 29 | 0 | 15 | 16 | 1 | 13 | 17 |
| 11 | 18 | 2.528 | 0 | 18 | 0 | 18 | 0 | 3 | 15 | 0 | 6 | 12 |
| 12 | 3 | 3.698 | 0 | 3 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 3 |

| G | Z | F | V5 | | | | V6 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 0 | 7 | -4.050 | 7 | 0 | 0 | 0 | 7 | 0 | 0 | 0 |
| 1 | 27 | -2.740 | 17 | 10 | 0 | 0 | 23 | 4 | 0 | 0 |
| 2 | 55 | -2.000 | 28 | 27 | 0 | 0 | 39 | 16 | 0 | 0 |
| 3 | 84 | -1.412 | 16 | 65 | 3 | 0 | 45 | 38 | 1 | 0 |
| 4 | 124 | -0.881 | 22 | 91 | 11 | 0 | 39 | 78 | 7 | 0 |
| 5 | 113 | -0.372 | 11 | 87 | 14 | 1 | 20 | 74 | 18 | 1 |
| 6 | 130 | 0.120 | 4 | 93 | 31 | 2 | 9 | 82 | 37 | 2 |
| 7 | 111 | 0.583 | 1 | 68 | 39 | 3 | 5 | 55 | 48 | 3 |
| 8 | 58 | 1.016 | 0 | 20 | 27 | 11 | 0 | 27 | 25 | 6 |
| 9 | 54 | 1.445 | 1 | 11 | 22 | 20 | 0 | 12 | 24 | 18 |
| 10 | 31 | 1.916 | 0 | 0 | 13 | 18 | 0 | 2 | 13 | 16 |
| 11 | 18 | 2.528 | 0 | 0 | 4 | 14 | 0 | 0 | 5 | 13 |
| 12 | 3 | 3.698 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 3 |

************Erläuterung**

Betrachten wir die 2. Zeile. Sie enthält die Personen, die G=1 Gesamtpunkte erzielten. Ihr Fähigkeitswert ist F= -2.740. Insgesamt haben Z=27 Personen 1 Gesamtpunkt erzielt. Der Item V1 wurde von 22 Personen mit 0 und von 5 Personen mit 1 beantwortet. Zusammen sind dies wieder Z=27 Personen. Entsprechendes gilt dann für V2 usw. Addiert man die Personen, die in V1 bis V6 mit 1 geantwortet haben, dann kommt man wieder auf Z=27.

### P14.8.8.3 Erläuterung zu den Grafiken bei dichotomen und polytomen Items

#### P14.8.8.3.1 Dichotome Items

Betrachten wir die Grafik für Item V2 aus den Daten „simRasch815.fre“. Zuvor gibt Almo folgende Tabelle aus:

Detailanalyse fuer Variable V2 Item_B

| Gesamtpunkt | Faehigkeitswert | Haeufigkeit gesamt | Haeufigkeit A2 | Ja-Wahrscheinlichkeit bei jeweiligem Gesamtpunkt empirisch | Ja-Wahrscheinlichkeit bei jeweiligem Gesamtpunkt theoretisch | Abweichung empirisch theoretisch |
|---|---|---|---|---|---|---|
| 0 x | -4.0501 | 7 | 0 | 0.0000 | 0.0406 | -0.0406 |
| 1 | -2.7402 | 27 | 1 | 0.0370 | 0.1356 | -0.0986 |
| 2 | -1.9997 | 55 | 10 | 0.1818 | 0.2476 | -0.0658 |
| 3 | -1.4117 | 84 | 26 | 0.3095 | 0.3720 | -0.0625 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 4 | | -0.8814 | 124 | 58 | 0.4677 | 0.5017 | -0.0339 |
| 5 | | -0.3723 | 113 | 79 | 0.6991 | 0.6262 | 0.0729 |
| **6** | | **0.1202** | **130** | **101** | **0.7769** | **0.7327** | **0.0442** |
| 7 | | 0.5826 | 111 | 97 | 0.8739 | 0.8132 | 0.0607 |
| 8 | | 1.0158 | 58 | 52 | 0.8966 | 0.8703 | 0.0262 |
| 9 | | 1.4449 | 54 | 48 | 0.8889 | 0.9116 | -0.0227 |
| 10 | | 1.9161 | 31 | 29 | 0.9355 | 0.9429 | -0.0074 |
| 11 | | 2.5276 | 18 | 18 | 1.0000 | 0.9682 | 0.0318 |
| 12 | x | 3.6979 | 3 | 3 | 1.0000 | 0.9899 | 0.0101 |

Mit "Ja-Wahrscheinlichkeit" ist die Wahrscheinlichkeit gemeint,
dass eine Person mit dem Gesamtpunkt G, das Item V2 positiv beantwortet

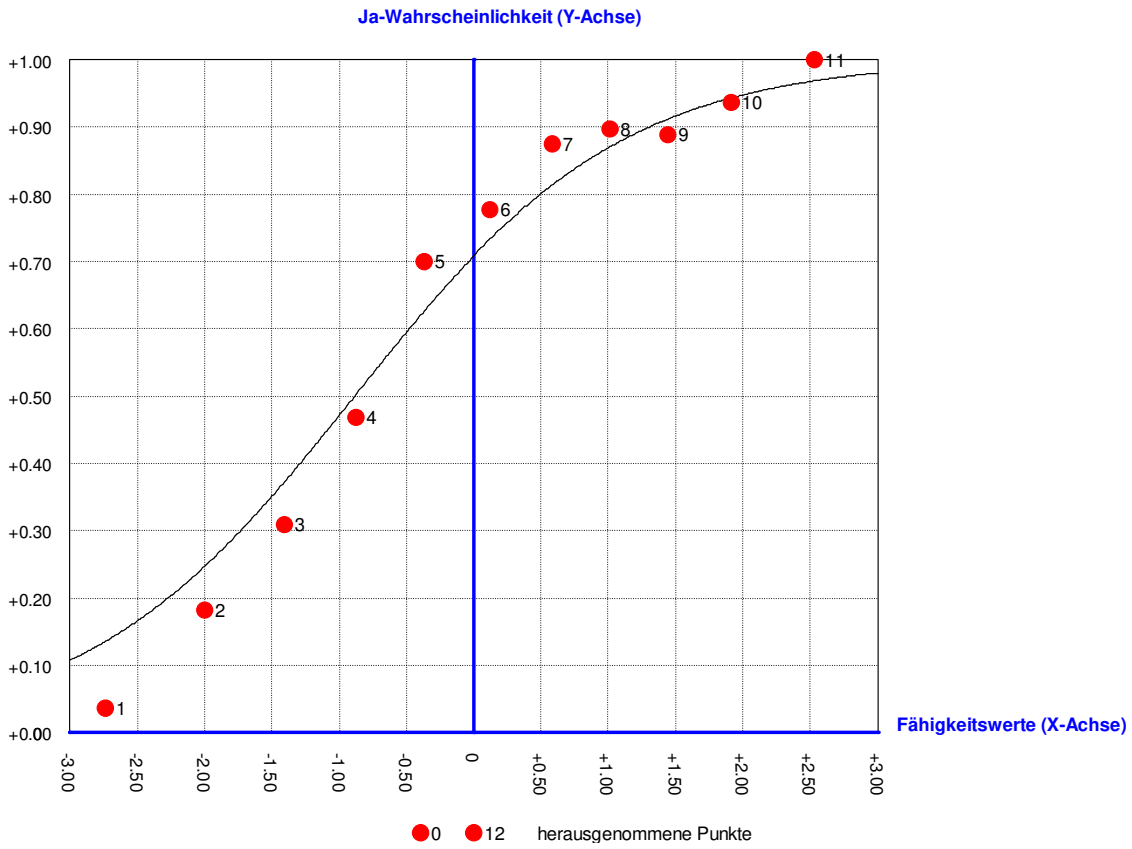
Mit "x" markierte Gesamtpunkte sollten nicht interpretiert werden
Die Punkte 0 und 12 haben notwendigerweise eine empirische "Ja-Wahrscheinlichkeit" von 0.0 bzw. 1.0. Sie werden nicht in die nachfolgende Grafik aufgenommen

"Ja-Wahrscheinlichkeit" bedeutet die Wahrscheinlichkeit auf den jeweiligen Item eine positiven Antwort zu geben. Bei einem Wissentest könnte das sein "hat richtig beantwortet". Bei einem Mathematiktest könnte das sein "hat Aufgabe richtig gelöst". Bei einer sozialwissenschaftlichen Meinungsfrage könnte das sein "hat Ja geantwortet".

Betrachten wir die Zeile mit 6 Gesamtpunkten.
Personen, die aus der Beantwortung aller Items 6 Gesamtpunkte erzielt haben besitzen einen Fähigkeitswert von 0.1202. Insgesamt haben 130 Personen 6 Gesamtpunkte erzielt. 101 von ihnen haben Item V2 positiv beantwortet; das sind 77.69% . Ihre empirische Wahrscheinlichkeit ist also 0.7769. Das dichotome Rasch-Modell prognostiziert für sie hingegen eine theoretische Wahrscheinlichkeit von 0.7327. Der Unterschied ist gering. Besonders deutlich wird das, wenn man die Grafik – und dort Punkt 6 – betrachtet.

Der kleinste Gesamtpunktwert 0 und der maximale Gesamtpunktwert 12 weisen eine Besonderheit auf. Beim kleinsten Gesamtpunkt 0 wurde keines der Items bejaht. Die empirische Wahrscheinlichkeit muss deswegen 0.0 sein. Beim maximalen Gesamtpunkt wurden alle 12 Items in der höchsten Kategorie bejaht. Demzufolge muss die empirische Wahrscheinlichkeit 1.0 sein. Die Punkte 0 und 12 werden deswegen aus der Grafik herausgenommen und zur Orientierung für den Benutzer unterhalb der Grafik platziert.

```
Logistische Funktion Y = 1/(1+e**(-0.888-X)) für Variable V2 Item_B
rote Punkte=empirische Ja-Wahrscheinlichkeit bei einem Gesamtpunktwert von x
Die Ges.punktwerte stehen bei den Punkten. Ges.punktwerte in Klammern geschätzt
```




Die Punkte liegen alle sehr dicht bei der theoretischen Kurve. Das Item V2 ist „Rasch-konform“.

### P14.8.8.3.2 Polytome Items

Betrachten wir die Grafiken für Item 3. Dieser Item besitzt 3 Ausprägungen. Almo zeichnet 3 Grafiken. Die Kategorienkurven in den 3 Grafiken sind dieselben. Unterschiedlich sind die eingezeichneten Gesamtpunkte. Betrachten wir die 2. Grafik für die Ausprägung 1. Almo gibt zuerst folgende Tabelle aus

```
Variable V3  Item_C:  Detailanalyse fuer Auspraegung A1
                                                  Ja-Wahrscheinlichkeit     Abweichung
                                Haeufigkeit   bei jeweiligem Gesamtpunkt   empirisch
Gesamtpunkt   Faehigkeitswert   gesamt   A1   empirisch      theoretisch  theoretisch
     0   x       -4.0501           7      0     0.0000          0.0444      -0.0444
     1           -2.7402          27      3     0.1111          0.1466      -0.0355
     2           -1.9997          55     16     0.2909          0.2628       0.0282
     3           -1.4117          84     29     0.3452          0.3837      -0.0385
     4           -0.8814         124     59     0.4758          0.4947      -0.0188
     5           -0.3723         113     69     0.6106          0.5752       0.0354
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **6** | | **0.1202** | **130** | **94** | **0.7231** | **0.6056** | **0.1174** |
| 7 | | 0.5826 | 111 | 69 | 0.6216 | 0.5829 | 0.0387 |
| 8 | | 1.0158 | 58 | 31 | 0.5345 | 0.5218 | 0.0127 |
| 9 | | 1.4449 | 54 | 22 | 0.4074 | 0.4371 | -0.0297 |
| 10 | | 1.9161 | 31 | 15 | 0.4839 | 0.3360 | 0.1479 |
| 11 | | 2.5276 | 18 | 3 | 0.1667 | 0.2188 | -0.0521 |
| 12 | x | 3.6979 | 3 | 0 | 0.0000 | 0.0805 | -0.0805 |

```
Mit "Ja-Wahrscheinlichkeit" ist die Wahrscheinlichkeit gemeint,
dass eine Person mit dem Gesamtpunkt G, aus den alternativen Kategorien
des Items V3 die Kategorie A1 waehlt

Mit  "x"   markierte Gesamtpunkte duerfen nicht interpretiert werden
aus folgendem Grund:
      1. Wenn Personen dieses Item mit 1 beantworten
         dann können Gesamtpunkte kleinergleich 0 nicht auftreten
      2. und dann können auch Gesamtpunkte grössergleich 12 nicht auftreten
Die mit "x" markierten Punkte werden nicht in die nachfolgende Grafik aufgenommen
```

"Ja-Wahrscheinlichkeit" bedeutet die Wahrscheinlichkeit aus den alternativen Kategorien des polytomen Items (im Beispiel: V3) die Kategorie x (im Beispiel: A1) gewählt zu haben

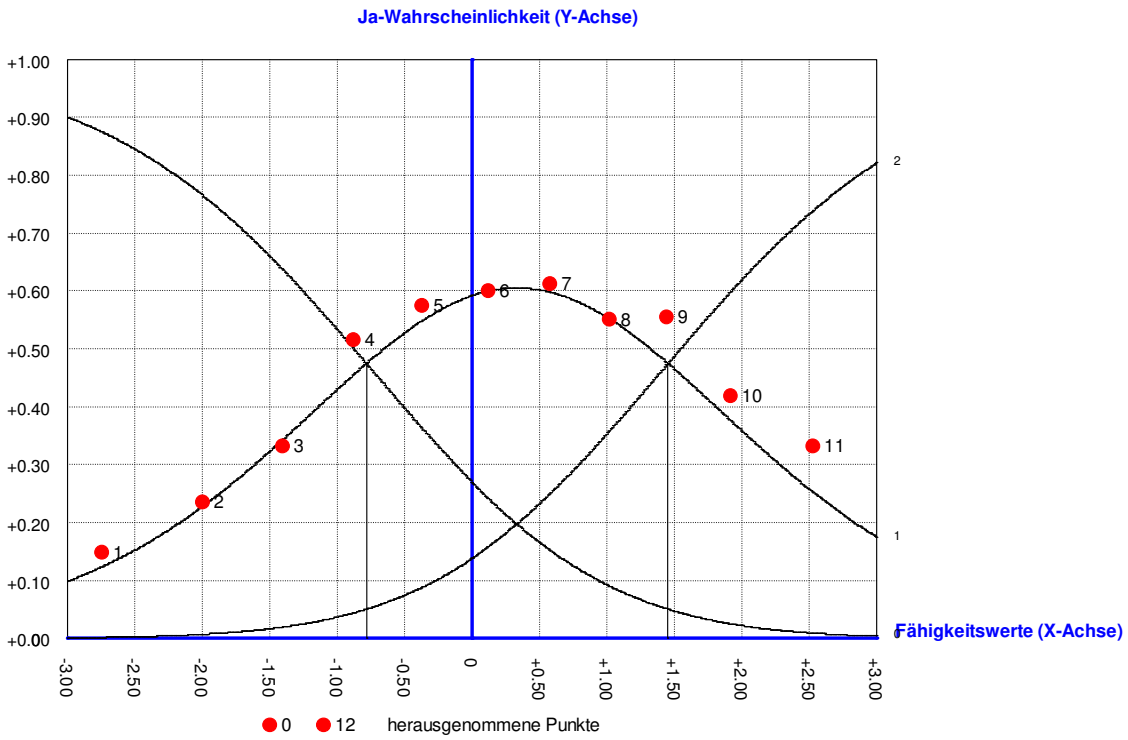
Beispiel:
Der Item 3 hat die 3 Ausprägungen

A0 leicht
A1 mittel
A2 schwer

Obige Tabelle bezieht sich auf die Kategorie A1 "mittel".

Betrachten wir Personen, die aus der Beantwortung aller Items 6 Gesamtpunkte erzielt haben. Sie besitzen einen Fähigkeitswert von 0.1202. Insgesamt haben 130 Personen einen Gesamtpunkt von 130 erzielt. 94 von ihnen haben im Item V3 die Ausprägung A1 gewählt; das sind 72.31% Ihre empirische Wahrscheinlichkeit ist also 0.7231. Das ordinale Rasch-Modell prognostiziert für sie hingegen eine theoretische Wahrscheinlichkeit von 0.6057. Die Grafik zu dieser Tabelle ist dann:.

```
Kategorienfunktionen fuer V4 Item_D mit Gesamtpunkten fuer Auspraegung 1
Auspraegung 0: gruene  Linie:
Auspraegung 1: blaue   Linie:
Auspraegung 2: gelbe   Linie:
```




```
Schwellenbereich der Auspraegung 0: von minus unendlich bis -0.9818
Schwellenbereich der Auspraegung 1: von -0.9818 bis  1.2630.   Breite: 2.2448
Schwellenbereich der Auspraegung 2: von  1.2630 bis plus unendlich

Beachte: Wenn Personen dieses Item mit 1 beantworten
         dann können Gesamtpunkte kleinergleich 0 nicht auftreten
Beachte: Wenn Personen dieses Item mit 1 beantworten
         dann können Gesamtpunkte grössergleich 12 nicht auftreten
Diese Punkte werden aus der Grafik herausgenommen und unterhalb der Grafik platziert
```

Kurve 2 ist eine logistische Kurve wie bei den dichotomen Items. Kurve 0 ist eine "umgedrehte" logistische Kurve. Eine Besonderheit ist die Kurve 1. Sie besitzt ein Maximum (kleiner 1) und läuft dann wieder nach unten.

Die 3 Kategorienfunktionen dieses Items weisen folgende Merkmale auf. Die eingezeichneten Punkte werden anschliessend diskutiert.

```
1. Für Personen mit Fähigkeitswerten von -unendlich bis -0.98 ist die
   Wahrscheinlichkeit, dass sie von den 3 alternativen Kategorien
   des Items die Kategorie 0 wählen am grössten. Von Fähigkeit -0.98
   bis +1.26 ist die Wahrscheinlichkeit für die Ausprägung 1 am grössten
   und von +1.26 bis plus unendlich die für Ausprägung 2.
   Die 2 Punkte -0.98 und +1.26 werden "Schwellenpunkte" genannt.
   Das Kontinuum von minus bis plus unendlich wird in 3 Bereiche
   untergliedert. Jede Kategorie besitzt ihren Bereich, in dem sie
   die dominierende Wahrscheinlichkeit besitzt.
```

```
2. Über jedem Punkt der x-Achse addieren sich die y-Werte, d.h.
   die Wahrscheinlichkeiten der 3 Kurven zu 1.0
```

Auch bei polytomen Items treten bei den minimalen und maximalen Gesamtpunkten Besonderheiten auf. Bei 0 Gesamtpunkten ist kein Item bejaht worden. Die empirische Wahrscheinlichkeit muss also 0.0 sein. Beim maximalen Gesamtpunkt 12 sind alle Items in der höchsten Kategorie bejaht worden. Personen, die im Item V3 die mittlere Kategorie, also nicht die höchste Kategorie bejaht haben, können nicht die maximale Zahl 12 der Gesamtpunkte erreichen. Wenn sie alle anderen Items mit der höchsten Kategorie bejahen, dann erzielen sie maximal 11 Punkte. Almo nimmt diese Punkte aus der Grafik heraus und platziert sie zur Orientierung für den Benutzer unterhalb der Grafik

**Die in die Grafik eingezeichneten Punkte.**
Dies sind Gesamtpunktwerte. In obiger Grafik wurden jene Personen herausgegriffen, die die Kategorie 1 gewählt haben. Betrachten wir Punkt 3. Insgesamt haben 84 Personen 3 Gesamtpunkte erzielt.
Davon haben sich beim betrachteten Item 29 für Kategorie 1 entschieden. Das ist ein Anteil von 29/84=0.35. Das ist ihre tatsächliche empirische Wahrscheinlichkeit die Ausprägung 1 zu wählen. Der entsprechende Punkt auf der Kurve ist die vom Rasch-Modell prognostizierte "theoretische Wahrscheinlichkeit" von 0.38. Die Gesamtpunkte liegen sehr dicht um die Kurve herum - vielleicht mit Ausnahme von Punkt 6.

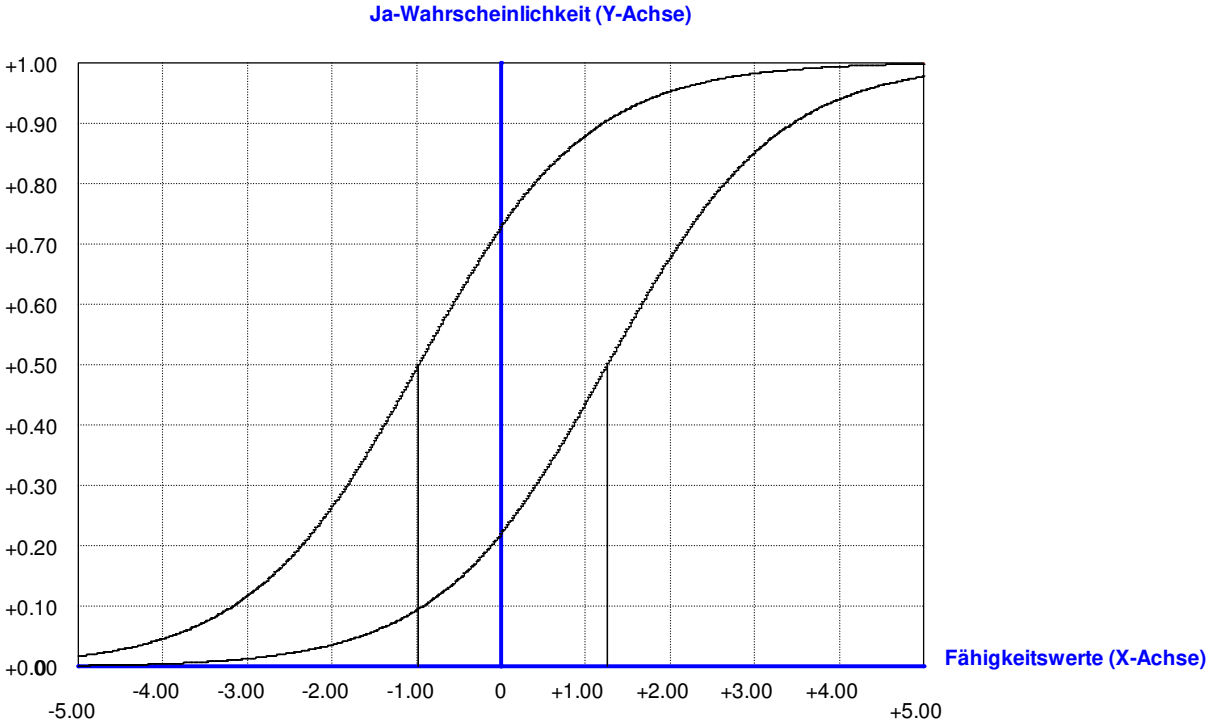
Entsprechende Grafiken zu Item 3 werden dann von Almo noch zu den
Ausprägungen 0 und 2 gezeichnet, also die Grafiken, in denen die Gesamtpunkte der Personen eingezeichnet sind, die sich für Ausprägung 0 und 2 entschieden haben.

So entstehen für dieses Item insgesamt 3 Grafiken. Das könnte auch in einer einzigen Grafik gemacht werden (mit verschiedenen Farben) - würde aber unübersichtlich werden.

In den drei Kurven liegen die Punkte alle sehr dicht bei den Kurven, so dass gefolgert werden kann: Das Item ist "Rasch-konform". Das ist natürlich ein intuitiver, dafür aber sehr anschaulicher Itemfit-Test. Mit den Koeffizienten "Outfit" und "Infit" werden in Almo standardisierte Tests angeboten, die allerdings ihre Probleme haben

**Kurve der Schwellen-Wahrscheinlichkeiten**

Als 4. Grafik zu Item 3 werden dann die "Schwellen-Wahrscheinlichkeiten" abgebildet. J. Rost verwendet diesen Begriff. Auch der Begriff "kumulative Wahrscheinlichkeit" wird verwendet.




Schwellen-Wahrscheinlichkeiten fuer V3 Item_C
Kurve 1: Übergangs-Wahrscheinlichkeit von Auspraegung 0 zu 1
Kurve 2: Übergangs-Wahrscheinlichkeit von Auspraegung 1 zu 2

Kurve 1: Uebergangs-Wahrscheinlichkeit von Katekorie 0 zu 1

| Gesamtpunktwert | Skalenwert Faehigkeitswert | Uebergangs-Wahrscheinlichkeit |
|---|---|---|
| 0 | -4.0501 | 0.0444 |
| 1 | -2.7402 | 0.1470 |
| **2** | **-1.9997** | **0.2654** |
| 3 | -1.4117 | 0.3941 |
| 4 | -0.8814 | 0.5251 |
| 5 | -0.3723 | 0.6478 |
| **6** | **0.1202** | **0.7506** |
| 7 | 0.5826 | 0.8270 |
| 8 | 1.0158 | 0.8805 |
| 9 | 1.4449 | 0.9188 |
| 10 | 1.9161 | 0.9477 |
| 11 | 2.5276 | 0.9710 |
| 12 | 3.6979 | 0.9908 |

Betrachten wir eine Person mit 2 Gesamtpunkten. Ihr Fähigkeitswert ist, wie man der Tabelle entnehmen kann, x= -1.9997. Diese Person liegt also im "Schwellenbereich minus unendlich bis -0.9818. Der y-Wert von Kurve 1 ist an dieser x-Stelle 0.2654. Das bedeutet: Diese Person hat eine Wahrscheinlichkeit von 0.2654 von der Ausprägung 0 zur Ausprägung 1 überzugehen - also eine sehr kleine Wahrscheinlichkeit. Sie wird wohl Ausprägung 0 wählen.

Eine Person mit dem höheren Gesamtpunkt 6 und dem dazugehörigen Fähigkeitswert 0.1202 hat in Kurve 1 einen y-Wert von 0.7506. Sie liegt damit im Schwellenbereich von -0.9818 bis 1.2630. Die Person besitzt damit eine Wahrscheinlichkeit von 0.7506 von der Ausprägung 0 zur Ausprägung 1 überzugehen.

Es kann also kurz definiert werden: "Schwellenwahrscheinlichkeit" ist die Wahrscheinlichkeit des Übergangs zur nächsten benachbarten Kategorie.

### P14.8.8.4 Itemfit-Masse

#### P14.8.8.4.1 Infit und Outfit

| | Chi-Quad. | Chi(norm) | outfit | t *) | p *) | infit | t *) | p *) |
|---|---|---|---|---|---|---|---|---|
| V1 | 687.72941 | 823.81607 | 1.01082 | 0.13416 | 0.89331 | 1.05494 | 1.08456 | 0.27844 |
| V2 | 711.86435 | 847.95100 | 1.04043 | 0.56382 | 0.57303 | 1.05052 | 1.38552 | 0.16627 |
| V3 | 746.73914 | 882.82580 | 1.08322 | 1.39354 | 0.16384 | 1.06651 | 1.50209 | 0.13346 |
| V4 | 739.22166 | 875.30832 | 1.07400 | 1.22349 | 0.22150 | 1.06181 | 1.38993 | 0.16493 |
| V5 | 589.36376 | 725.45041 | 0.89012 | -1.78198 | 0.07512 | 0.87626 | -2.55834 | 0.01070 |
| V6 | 598.56174 | 734.64840 | 0.90141 | -1.66571 | 0.09616 | 0.88995 | -2.36999 | 0.01802 |

Freiheitsgrade fuer t: 815

*) ein t-Wert zwischen -1.96 und +1.96, bzw. ein p-Wert zwischen 0.05 und 0.95
bedeutet: die Variable ist "Rasch-konform"

Outfit und Infit geben an, wie gut ein Item der Annahmen der gleichen Trennschärfen des Rasch-Modells entspricht. Beide beruhen auf quadrierten standardisierten Residuen, deren Summe je Item zunächst einen Chi-Quadrat-Wert ergibt. Dieser wird normiert nach der Formel

```
diff       = ChiQuad - ChiMittelwert aus allen m Items
ChiQuad(norm) = N + diff

m=Itemzahl N=Personenzahl.
```

Outfit ist dann

```
Outfit = ChiQuad(norm)/N.
```

Der normierte Chi-Quadrat-Wert wird also nochmals normiert.

Die etwas seltsame Folge ist, dass wenn die (nichtnormierten) Chi-Quadrat-Werte aller Items gleich sind, die ChiQuad(norm)-Werte =N und Outfit =1 ist - unabhängig davon wie gross die (nichtnormierten) ChiQuad-Werte sind. Outfit und Infit sind somit relative Masse. Sie vergleichen die Trennschärfen der Items untereinander.

Weicht jedoch ein Item in seinem (nichtnormierten) ChiQuad-Wert stark vom Mittelwert ab, dann ist das ein Hinweis darauf, dass es nicht zu den anderen dazugehört. Beim dichotomen Item kann man dann folgern, dass u.a. seine Itemkurve in Wirklichkeit flacher oder steiler verläuft als die der anderen. Das Rasch-Modell unterstellt, dass an alle Items dieselbe logistische Funktion mit derselben Steilheit angepasst werden kann. Nur der Parameter, der die horizontale Lage der Funktion bestimmt (=der Schwierigkeitsparameter) darf variieren. Die Steilheit der logistischen Kurve kann als Trennschärfe interpretiert werden. Dies ist die Fähigkeit des Items die Personen zu trennen. Outfit und Infit dürfen jedoch nicht als Maße angesehen werden, die die Trennschärfe ausdrücken. Sie dienen dazu, Items zu identifizieren, die "anders sind als die anderen", die eine andere Trennschärfe besitzen als die anderen. Die Ursache für ein in der Trennschärfe "abweichendes" Item kann sein, dass das Item auf einer anderen Dimension misst als die anderen

oder dass es zusätzlich noch auf einer 2. oder weiteren Dimensionen misst.

Bei polytomen Items kann die Bedeutung von ChiQuad(norm) und Outfit nicht so anschaulich gemacht werden.

Der Idealzustand ist gegeben, wenn `ChiQuad(norm)=N` bzw. `Outfit=1` ist Dann ist das Item "Rasch-konform".

Infit ist in der gleichen Weise zu interpretieren. Seine Berechnung ist jedoch komplexer. Das C-Unterprogramm zu Outfit und Infit ist in Almo in den "Algorithmen_in_C" im Modul a_up_algorith11.c enthalten oder in „Rasch_in_R“ in der R-Funktion „Itemfit“.

Wright & Linacre geben folgende Interpretation von In- und Outfit: Der Wert von Outfit und Infit reicht von 0 über 1 bis unendlich. Ein Wert grösser 1, z.B. 1.2 bedeutet, dass 20% mehr "Zufälligkeit" in den Daten des betreffenden Items enthalten ist, als das Rasch-Modell annimmt. Ein Wert kleiner 1, z.B. 0.7 bedeutet, dass das betreffende Item zu sehr einem "Guttman-Item" gleicht. Bei einem dichotomen Item bedeutet dies, dass die Antworten der Personen nicht der S-förmigen logistischen Kurve entsprechen (die das Rasch-Modell unterstellt und den Daten aufzwingt). Sie entsprechen vielmehr dem stufenförmigen Verlauf des Guttman-Verfahrens. Wright & Linacre empfehlen Items mit einem Outfit- oder Infit-Wert zwischen 0.5 und 1.5 zu akzeptieren. Items über 2 sollten unbedingt ausgeschlossen werden.

Outfit und Infit können dann noch mit der Wilson-Hilferty-Transformation zu einem z- bzw. t-Wert transformiert werden. Damit ist dann die Aussage möglich.

```
Ein Item ist Rasch-konform,
wenn sein t-Wert aus Outfit bzw. Infit
zwischen -1.96 und +1.96 liegt
```

Diese Regel ist strenger als die Empfehlungen von Wright & Linacre

Wie Potuschak in P14.8.9.6.2 **a**usführt sind die t-Werte aus infit und outfit nicht standard-normaverteilt. Das hat zur Folge, dass eigentlich keine p-Werte berechnet werden dürfen und dass die t-Werte schlicht als Indikatoren des Itemfits begriffen werden müssen.

**Literatur**:

Wright, Masters: Rating Scale Analysis, 1982, in www.rasch.org/rmt/rmt34e.htm
Jürgen Rost: Lehrbuch Testtheorie-Testkonstruktion, 2004, Kap. 6.2
B. Wright & M. Linacre: Reasonable mean-square fit values, www.rasch.org/rmt/rmt83b.htm

### P14.8.8.4.2 Q-Index (nach Rost/Davier)

Almo liefert für die Testdaten "simRasch815.fre" folgende Ausgabe:

| Variable | Q-Index |
|---|---|
| V1 | 0.180968 |
| V2 | 0.164744 |
| V3 | 0.161071 |
| V4 | 0.163420 |
| V5 | 0.108644 |
| V6 | 0.105813 |

Ein Q-Index von .0 bedeutet, dass die Kurve des Items eine "Sprungkurve" wie im Guttman-Modell ist. Die Trennschärfe des Items ist maximal. Ein Q-Index von 0.5 bedeutet, dass die Kurve des Items eine horizontal verlaufende Gerade ist. Das Item kann Personen unterschiedlicher Fähigkeit nicht unterscheiden. Werte über 0.5 bedeuten, dass das Items eine "umgedrehte" Trennschärfe besitzt Die Itemkurve verläuft von links oben nach rechts unten. Je schwieriger das Item ist umso eher wird es mit abnehmender Fähigkeit gelöst. Bei 1.0 besitzt das Item eine "umgedrehte" Sprungfunktion. "Brauchbare" Items müssen also einen Wert zwischen .0 und 0.5 besitzen. Ab welchem Wert im Bereich zwischen 0 und 0.5 ein Item nicht mehr akzeptabel ist, kann nicht entschieden werden.

Der Q-Index ist ein Indikator für den Trennschärfe-Parameter der Items. Er ist nur ein Indikator und nicht selbst der Trennschärfe-Koeffizient. Je kleiner der Wert umso steiler verläuft (bei dichotomen Items) die logistische Kurve; und je grösser er ist, umso flacher verläuft die Kurve.

Davier bietet in seinem Programm „Winmira“ dann noch einen z-Wert und p-Wert an, der überprüft, ob die Items dieselbe Trennschärfe besitzen. Es wird nicht überprüft, ob die Trennschärfe der Items in einem akzeptablen Bereich liegt. Bei einer Simulation mit 5 Items mit der extremen Trennschärfe c=16 und 1 Item mit der normalen Trennschärfe c=1 wird letzterer durch den z- bzw. p-Wert als abweichend erkannt und als „nicht-Rasch-konform“ markiert.

Wo sollte der Q-Index liegen, damit das Item als brauchbar akzeptiert werden kann. Der Idealwert liegt nicht in der Mitte zwischen 0 und 0.5. Er ist auch abhängig von der Zahl der Kategorien des Items und den anderen Parametern des Rasch-Modells. Insofern sind die Q-Indices verschiedener Items (zumindest unterschiedlicher Kategorienzahl) nicht miteinander vergleichbar.

Almo errechnet, wenn der Benutzer dies in der Optionsbox "Itemfit" anfordert, für jedes Item einen Idealwert. Der von Almo errechnete Idealwert besitzt einen Trennschschärfe-Koeffizienten von c=1.0. Dem Benutzer wird es damit ermöglicht, abzuschätzen, wie weit der durch den Q-Index ausgedrückte Trennschärfe-Koeffizient von 1.0 entfernt ist. Es muss aber festgehalten werden, dass das Rasch-Modell nur gleiche Trennschärfe für alle Items fordert. Zulässig wäre also auch ein für alle Items gleicher Trennschärfe-Koeffizient von beispielsweise c=2.0. Siehe dazu den nachfolgenden Abschnitt P14.8.8.4.2.1 „Die Steigung der logistischen Rasch-Funktion“.

Der Idealwert kann durch 2 Methoden ermittelt werden.

**Methode 1: Q-Index-Idealwert durch "Rasch-konforme" Zufallsdaten.**

Almo simuliert (wie beim bootstrapping für den pauschalen Modelltest) mit Zufallszahlen eine Rasch-konforme Datei - wobei aus den empirischen Daten folgende Parameter für die Simulation vorgegeben werden:

Zahl der Items,
Zahl der Ausprägungen je Item,
Schwierigkeitsparameter,
Fähigkeitsparameter,
Personenzahl je Gesamtpunkt,

Bei der Simulation wird der Trennschärfeparameter auf 1 gesetzt. Siehe dazu nachfolgend den

Abschnitt P14.8.8.4.2.1. Siehe auch das C-Programm "simord" oder „simord2“ in den „Algorithmen_in_C“ im Modul up_algorith11.c oder das entsprechende R-Programm im Ordner „Rasch_in_R“.
Es wird also eine N*m-Datei erzeugt, die der empirischen Datei äquivalent ist – mit dem entscheidenden Unterschied, dass die simulierten Daten "Rasch-konform" sind.

Aus dieser Datei wird der Q-Index je Item errechnet. Dieser Vorgang wird 200 mal wiederholt, d.h. es werden 200 Dateien erzeugt und damit 200 Q-Indices für jeden einzelnen Item. Danach wird der Mittelwert aus diesen 200 Q-Indices je Item ermittelt. Dies ist dann der "Idealwert". Der Vorgang ist zeitintensiv. Je grösser die Itemzahl, die Personenzahl und die Kategorienzahl der Items umso länger die Rechenzeit. In der Regel beträgt die Rechenzeit aber nur wenige Sekunden.

Almo gibt dann noch das Intervall von plus/minus 2 Standardabweichungen um den Idealwert an und markiert Items, die ausserhalb dieses Intervalls liegen. Dies ist kein statistischer Test. Der Benutzer erhält nur eine Entscheidungshilfe. Wichtig ist auch zu bedenken, dass bei der Simulation des Idealwerts ein „Trennschärfeparameter“ von 1 verwendet wurde. Durch den Vergleich von Idealwert und tatsächlichem Q-Wert wird also überprüft, wie weit sich die Trennschärfe des betreffenden Items von c=1 entfernt. Siehe dazu P14.8.8.4.2.1 „Die Steigung der logistischen Rasch-Funktion“.

Almo liefert für die Testdaten "simRasch815.fre" folgende Ausgabe:

```
Itemfit: Q-Index (nach Rost/von Davier)
---------------------------------------
                    Idealwert
                    aus 200                      Idealwert       Idealwert
Variable  Q-Index   Simulationen Stand.abweichg  minus 2*Stdabwg plus 2*Stdabwg
--------  -------   ------------ --------------  --------------- --------------
   V1     0.18097    0.15911      0.0143227796       0.13046        0.18775
   V2     0.16474    0.15376      0.0142163618       0.12533        0.18220
   V3     0.16107    0.11888      0.0087454553       0.10139        0.13637   ?
   V4     0.16342    0.12030      0.0087548335       0.10279        0.13781   ?
   V5     0.10864    0.09929      0.0079173263       0.08346        0.11513
   V6     0.10581    0.09501      0.0066371379       0.08174        0.10829
```

Die beiden Items V3 und V4 liegen ausserhalb des Intervalls Idealwert plus/minus 2 Standardabweichungen. Sie werden durch ein Fragezeichen markiert.

**Methode 2: Q-Index-Idealwert durch "Rasch-perfekte" Daten.**

Almo verwendet aus den empirischen Daten folgende Werte

Personenzahl
Zahl der Items
Zahl der Ausprägungen der Items
Zahl aller möglichen Antwortmuster
Zahl der Personen je Gesamtpunkt
Schwierigkeitsparameter je Item und Ausprägung
Fähigkeitswerte je Gesamtpunkt

Bei der Simulation wird der Trennschärfeparameter auf 1 gesetzt. Siehe dazu nachfolgend den

Abschnitt P14.8.8.4.2.1. Siehe auch das C-Programm "simperfekt" in den „Algorithmen_in_C“ im Modul up_algorith11.c oder das entsprechende R-Programm „SIMexakt“ im Ordner „Rasch_in_R“.

Damit werden alle möglichen Antwortmuster erzeugt und je Antwortmuster artifizielle Personen. Dabei werden die Rasch-Formeln (0a) bzw. (1) verwendet. Insgesamt generiert Almo mindestens 10000 Personen.
Auf diese Weise entstehen perfekte Rasch-Daten. Siehe dazu das Unterprogramm "simPerfekt" und "Q_Index_Idealwert_perfekt" in den Algorithmen_in_C Modul up_algorith11. Wenn die generierten Daten „perfekt“ sind, dann müssen ihre Schwierigkeitswerte exakt gleich sein mit denen, die zu ihrer Erzeugung eingesetzt wurden. Es gibt aber ein Problem. Rasch-Daten müssen ganzzahlig sein, bei dichotmen Items 0, 1, bei polytomen 0, 1, 2, .... Das Programm „simPerfekt“ hingegen errechnet Lösungswahrscheinlichkeiten p, die fast immer zwischen den Ganzzahlen liegen. Im Programm muss deswegen p auf eine Ganzzahl gerundet werden. Dadurch können erhebliche Verzerrungen entstehen - so dass die Daten keineswegs mehr "Rasch-perfekt" sind.

Wird in der Box "Itemfit" im 3. Eingabefeld eine 3 eingegeben, dann überlässt der Benutzer Almo die Wahl, welche der beiden Methoden angewendet wird. Almo verfährt dann so: Ist die Zahl der Personen kleiner als die Zahl aller möglichen Antwortmuster, dann wird für die Methode der "Rasch-konformen" Zufallsdaten entschieden (bei denen dieses Rundungsproblem nicht entsteht).

Almo liefert für die Testdaten "simRasch815.fre" folgende Ausgabe:

```
Itemfit: Q-Index (nach Rost/von Davier)
---------------------------------------
                 "Rasch-perfekter" 20% Abweichung Idealwert        Idealwert
Variable Q-Index  Idealwert        von Idealwert  minus Abweichung plus Abweichung
-------- ------- ----------------- -------------- ---------------- ---------------
   V1    0.18097    0.15334        0.0306674244        0.12267       0.18400
   V2    0.16474    0.15500        0.0310006419        0.12400       0.18600
   V3    0.16107    0.11814        0.0236278911        0.09451       0.14177    ?
   V4    0.16342    0.12154        0.0243076206        0.09723       0.14585    ?
   V5    0.10864    0.09874        0.0197489074        0.07900       0.11849
   V6    0.10581    0.09467        0.0189345815        0.07574       0.11361
```

Die beiden Items V3 und V4 liegen ausserhalb des Intervalls Idealwert plus/minus 20% Abweichung. Sie werden durch ein Fragezeichen markiert.

Die Idealwerte der beiden Methoden sind in unserem Beispiel erst an der 3. oder 4. Kommastelle verschieden. Es sei nochmals darauf hingewiesen, dass beide Idealwert-Methoden keine statistischen Tests sind. Sie sollen dem Benutzer nur eine Orientierungshilfe für die Identifizierung unbrauchbarer Items geben. Wichtig ist auch zu bedenken, dass der Berechnung des Idealwerts ein „Trennschärfeparameter“ von c=1 verwendet wurde.

**Literatur zu Q-Index:**
Jürgen Rost, Matthias von Davier: A conditional item-fit index for Rasch-models, Applied psychological measurement, Vol. 18, June 1994, S.171-182

### P14.8.8.4.2.1 Die Steigung der logistischen Rasch-Funktion

Zu diesem Thema siehe auch P14.8.7.10, Eingabefeld 4: Trennschärfe der Items.
Eine Bedingung des Rasch-Modells ist es, dass die Trennschärfen der Items gleich sind. Wird der "Trennschärfe-Parameter“ explizit in die die Formel für die dichtomen Items eingeschrieben, dann müsste korrekterweise die Formel so geschrieben werden:

$$p = \frac{1}{1+e^{(c*(s-f))}}$$

```
e  = e-Zahl 2.718...
p  = Wahrscheinlichkeit einer positiven Antwort
s  = Schwierigkeitsparameter
f  = Fähigkeitsparameter
c  = Steigungsparameter der logistischen Kurve
ap = Zahl der Item-Kategorien minus 1
```

c ist der Steigungsparameter der logistischen Kurve. c kann als "Trennschärfeparameter" bezeichnet werden. Die Steigung der Tangente im Wendepunkt der Kurve ist c. Im Rasch-Modell muss c für alle Items denselben Wert besitzen.

Anders im "Birnbaum-Modell" (das die obige Formel verwendet), bei dem die Steigungen der logistischen Funktion verschieden sein dürfen - wodurch sich die Kurven überschneiden können. Siehe dazu Rost (2004, S. 133, 370).

Auch für polytome Items müsste die allgemeine Formel lauten:

$$p = \frac{e^{(c*(j*f-s))}}{sum}$$ für Itemkategorie j

wobei

$$sum = \sum_{k=0}^{ap} e^{(c*(k*f-s))}$$ Summierung von k=0 bis k=ap

Im Rasch-Modell muss c für alle Items denselben Wert besitzen. Üblicherweise wird deswegen in dieser Formel c weggelassen; d.h. dass c auf 1 gesetzt wird.

Das 2-parametrische „partial credit model“ nach Muraki (1992) verwendet diese allgemeine Formel und erlaubt, dass die c-Werte der Items verschieden sein dürfen.

Es sind sogar 3-parametrische Modell möglich. Deren Formel lautet für dichotome Items:

$$p = r + (1-r) * \frac{e^{(c*(f-s))}}{1+e^{(c*(f-s))}}$$

r wird als „Rateparameter“ bezeichnet, der zwischen 0 und 1 liegen kann. r muss für alle Items gleich sein. c darf verscgieden sein. Ein Anwendungsbeispiel wären etwa multiple choice Items, bei denen sich der Proband zwischen 2 vorgegebenen Antworten für eine entscheiden *muss*. Und wenn er ahnungslos ist, sich zufällig für eine entscheidet. Grafisch bedeutet dies, dass sich die S-förmige logistische Funktion asymptotisch einer Untergrenze von y=r annähert. Der Parameter r

muss vom Benutzer geschätzt werden. Siehe dazu Rost (2004, S.133).

In Almo werden vor allem folgende Koeffizienten verwendet um die (gleiche) Trennschärfe der Items im Rasch-Modell zu beurteilen:

Infit und Outfit
Q-Index (nach Rost/von Davier)

Für Infit und Outfit ist es egal wie groß c ist. Entscheidend ist, dass c für alle Items gleich ist. Infit und outfit reagieren nicht auf die Größe von c, wenn diese für alle Items gleich ist. Anders der Q-Index, der auf die Größe von c reagiert.

Wir haben 6 Rasch-konforme Zufallsdateien mit 6 verschiedenen Steigungparametern simuliert. Jede Datei enthält 20 dichotome Items. Alle Items einer Datei haben dieselbe Steigung c. Die Dateien für c=1, c=4 und c=8 sind unter dem Namen „Raschkonform1.fre“, „Raschkonform4.fre“, „Raschkonform8.fre“ im Almo-Ordner TESTDAT enthalten.

| c | Steigung der Tangente im Wendepunkt in Grad | Infit/Outfit | Q-Index | p |
|---|---|---|---|---|
| 0.25 | 14 | ca. 1.0 | ca. 0.330 | ca. 0.5 |
| 1 | 45 | ca. 1.0 | ca. 0.170 | ca. 0.5 |
| 4 | 76 | ca. 1.0 | ca. 0.033 | ca. 0.5 |
| 8 | 82 | ca. 1.0 | ca. 0.015 | ca. 0.5 |
| 16 | 86 | ca. 1.0 | ca. 0.006 | ca. 0.5 |
| 46 | 89 | ca. 1.0 | ca. 0.005 | ca. 0.5 |

Beispiel 1. Zeile: 20 Items mit Steigung c=0.25 wurden simuliert. Ihre Infit-/Outfit-Werte schwanken minimal um 1.0, ihr Q-Index minimal um 0.330.
Infit und Outfit bleiben in allen 6 Fällen "unbeeindruckt" – sogar bei einer extremen Steigung von 89 Grad, was im Prinzip einem Guttman-Item entspricht. Alle Infit-und Outfit-Werte variieren um 1.0 herum. Der Q-Index hingegen reagiert auf die zunehmende Steigung. Sein Wert nimmt ab. Er variiert geringfügig um den in der Tabelle angegeben ungefähren Wert. Der im Programm Winmira ausgegebene p-Wert von Q hingegen bleibt gleich. Er liegt mit ca. 0.5 ideal im "Rasch-konformen" Bereich von 0.05 bis 0.95. Er erkennt, dass die Steigungen aller Items gleich sind – jedoch nicht, ob die Steigung der Items in einem akzeptablen Bereich liegt.

Wird in Almo der "Idealwert" für den Q-Index simuliert, dann wird im Simulations-Kalkül eine Steigung c=1 unterstellt. Das ist zu bedenken, wenn der Wert von Q mit dem Idealwert verglichen wird um ein Urteil über die Brauchbarkeit des jeweiligen Items zu fällen.

Offen bleibt die Frage, wie flach oder wie steil die logistische Kurve verlaufen darf, damit das Rasch- Modell noch brauchbare Ergebnisse liefert.

#### P14.8.8.5 Modelltest und Bootstrap

Betrachten wir als Beispiel die Pisa-Studie 2009. Untersucht werden die 20 Mathematik-Testfragen des "Booklet 1" für 467 österreichische Schüler. 3 Items wurden zuvor herausgenommen. Das

Ergebnis ist folgendes

```
Zahl der Probanden                        467
Zahl aller moeglichen Antwortmuster    1048576
Zahl der besetzten Antwortmuster          455
Zahl der leeren Antwortmuster          1048121


******* von den 455 besetzten Antwortmuster sind 455 mit weniger als 5 Personen besetzt
******* von den 455 besetzten Antwortmuster sind 455 mit weniger als 5 Erwartungswerten
        besetzt
******* Wird der pauschale Modelltest gerechnet, dann entstehen
        valide Ergebnisse nur durch den Bootstrap-Test


      likelihood ratio    |     Pearsons Chi-Quadrat     |      Cressie-Read-Test
      LR      df     p    |  Chi-Quadrat     df     p    |  Chi-Quadrat     df     p
---------  -----   ------ | ------------  -----   ------ | ------------  -----   ------
  5111.07 1048536  1.0000 |   1102862.76 1048536  0.1349 |     94256.55 1048536  1.0000
```

Alle 3 p-Werte sind grösser als 0.05. Es wäre also zu interpretieren, dass das Modell gültig ist.
In aller Regel sind die p-Werte der 3 Tests verschieden, oft sogar erheblich verschieden. In der Literatur wird dem Cressie-Read-Test (CR-Test) der Vorzug gegeben. Wir würden auf Grund unserer Erfahrungen empfehlen prinzipiell nicht gegen den CR-Test zu entscheiden. Verwirft der CR-Test die Annahme, dass das Modell gültig ist (p kleiner 0.05) dann sollte man das akzeptieren. Almo testet das Modell gegen das saturierte Modell (siehe Rost, 2004, S. 334).

Die 467 Schüler verteilen sich auf 455 verschiedene Antwortnuster. 455 Antwortmuster sind mit weniger als 5 Personen besetzt. Ein Chi-Quadrat-Test darf nicht gerechnet werden. Eine Faustformel lautet, dass jede Zelle mit mindestens 5 Personen besetzt sein muss.

Es muss ein bootstrap-Test gerechnet werden, der auch bei dieser Datenlage nur 2 bis 3 Sekunden dauert

```
Bootstrap:
Es werden 1000 Bootstrap-Runden gerechnet. Davon werden 10 angezeigt


      likelihood ratio    |     Pearsons Chi-Quadrat     |      Cressie-Read-Test
      LR      df     p    |  Chi-Quadrat     df     p    |  Chi-Quadrat     df     p
---------  -----   ------ | ------------  -----   ------ | ------------  -----   ------
  4963.71 1048536  1.0000 |   1262915.69 1048536  0.0001 |     90367.30 1048536  1.0000
  4979.74 1048536  1.0000 |    912019.96 1048536  0.9991 |     83968.13 1048536  1.0000
  5021.79 1048536  1.0000 |   1030690.42 1048536  0.6555 |     90842.99 1048536  1.0000
  4984.05 1048536  1.0000 |    715356.80 1048536  1.0000 |     78890.88 1048536  1.0000
  4876.83 1048536  1.0000 |    891039.69 1048536  0.9998 |     81895.25 1048536  1.0000
  4915.72 1048536  1.0000 |    964473.65 1048536  0.9721 |     83379.41 1048536  1.0000
  4917.82 1048536  1.0000 |    975789.14 1048536  0.9498 |     84045.77 1048536  1.0000
  5026.12 1048536  1.0000 |    873571.73 1048536  1.0000 |     83472.80 1048536  1.0000
  5067.68 1048536  1.0000 |    846335.46 1048536  1.0000 |     84826.49 1048536  1.0000
-------------------------|------------------  ----------|----------------------------
der Testwert             |                               |
aus Bootstrap            |                               |
war hoeher als           |                               |
Modelltestwert           |                               |
in x Fällen          51  |          291                  |          191
Anteilswert p      0.05  |         0.29                  |         0.19

                    Ein p-Wert zwischen 0.05 und 0.95 bedeutet, dass die
```

```
                    Untersuchungsdaten "Rasch-konform" sind

                          |                              |
95 %-iges                 |                              |
Vertauens-                |                              |
intervall +/-       0.01  |         0.03                 |         0.02
```

Die Bedingung "p zwischen `0.05` und `0.95`" ist für alle 3 Test erfüllt.
Auch hier sind in der Regel die Ergebnisse aus den 3 Tests verschieden.

*Vergleich mit Winmira und Multira*
Die 3 pauschalen Modelltests liefern bei Almo, Winmira und Multira teilweise unterschiedliche aber sehr ähnliche Ergebnisse. Auch Winmira und Multira stimmen nicht exakt überein. Die Unterschiede müssen in etwas veränderten Berechnungsformeln liegen. Almo verwendet die nachfolgend in Abschnitt P14.8.9.6.1 angegebenen Formeln. Siehe auch das entsprechende C-Unterprogramm im Modul up_algorith11.c im Almo-Ordner „Algorithmrn_in_C“ bzw. das R-Programm im Almo-Ordner „Rasch_in_R“.

## Heinrich Potuschak

### P14.8.9 Parameterschätzungen und Modelltests im ordinalen Rasch-Modell

**Bezeichnungen**

m = Zahl der Items
$k_i$ Zahl der Kategorien, d.h. Ausprägungen mit x = 0, …, $k_i$-1 Stufen je Item i

mscs = $\sum_{i=1}^{m}(k_i - 1)$ = maximale Scoresumme,

amp = $\prod_{i=1}^{m} k_i$ = Anzahl aller möglichen Antwortmuster (pattern)

$np_r$ = Anzahl der Personen mit Scoresumme r, wobei
$0 \le r \le$ mscs und Summe($np_r$) = N = Anzahl der Personen.

$nt_{i,x}$ = Anzahl der Personen, die Item i bis zur Stufe x gelöst haben, wobei
$1 \le i \le m$, $1 \le x \le k_i$-1 und $nt_{i,0}$ = N – Summe($nt_{i,x}$).

$f_r$ = Fähigkeitsparameter einer Person mit Scoresumme r, wobei $0 \le r \le$ mscs.
Die fu maximieren die ungewichtete, die fw die gewichtete Likelihood.

$\tau$ = Schwellenparameter. Dimension = Dim(nt), d.h. m Zeilen mit je $k_i$-1 Spalten.
Jede Zeile ist aufsteigend geordnet, die Gesamtsumme ist 0. Def: $\tau_{i,0} = s_{i,0} = 0$.

s = Schwierigkeitsparameter, wobei $s_{i,x} = \sum_{y=1}^{x} \tau_{i,y} \Rightarrow \sum_{i=1}^{m} s(i, k_i - 1) = 0$ und $\tau_{i,x} = s_{i,x} - s_{i,x-1}$.

Die su maximieren die unbedingte, die sb die bedingte Likelihood.

su = Schwierigkeitsparameter aus unbedingter Schätzung (werden in Almo nicht ermittelt)
sb = Schwierigkeitsparameter aus bedingter Schätzung

pat = Ein einzelnes m-dimensionales Pattern.
pah = Vektor der Pattern-Häufigkeiten. Die Länge n0 ist die Zahl der besetzten Pattern.
dez = Dezimalcodes der n0 Pattern.

dezN = Dezimalcodes aller N Pattern.

### P14.8.9.1 Kodierung der Pattern

N, m und k seien bekannt. Die Basis des Zahlensystems, das ein Pattern in eine Dezimalzahl umwandelt, ist m-dimensional. Man beginnt mit $bas_m = 1$ und berechnet $bas_i = bas_{i+1}*k_{i+1}$ für i = m-1 bis1. Der Dezimalcode eines Patterns ist die Summe aller $pat_i*bas_i$ .
Die Rückrechnung beginnt mit $pat_1 = dez/bas_1$ und berechnet die restlichen $pat_i$ als Divisionsrest geteilt durch $bas_i$. Zusammen mit den pah ist es möglich, alle Daten anstelle einer N∗m- Matrix mit 2 Vektoren der Länge n0 zu speichern.

### P14.8.9.2 Schätzung der bedingten Schwierigkeitsparameter (sb)

Notwendige Bedingungen für eine reelle Lösung sind $nt_{i,0} > np_0$ , $nt_{i,k[i]-1} > np_{mscs+1}$ und $0 < nt_{i,x} < N$ für alle i und $0 \le x \le k_i-1$. Die Schätzung ist gültig, wenn das folgende Verfahren konvergiert und alle $\tau_{i,.}$ aufsteigend geordnet sind.
Die modellgemäße Wahrscheinlichkeit einer Person mit Fähigkeitsparameter $fw_r$ ,das Item i bis zur Stufe x zu lösen, lautet $P(i,x|r) = \exp[x*fw_r - sb_{i,x(i)}] / \sum_{y=0}^{k(i)-1} \exp[y*fw_r - sb_{i,y}]$ .

Die unbedingte Wahrscheinlichkeit, das Pattern pat zu beobachten, ist gleich dem Produkt aller $P(i,x_i|r)$. Die sb werden unter der Bedingung geschätzt, dass bei festgehaltenem r die Summe aller Patternwahrscheinlichkeiten gleich $np_r/N$ wird – daraus folgt die bedingte Patternwahrscheinlichkeit $P(\underline{x}|r) = \exp\left[-\sum_{i=1}^{m} sb_{i,x(i)}\right] / \gamma_r$ .

Die conditional loglikelihood lautet $cLL(sb) = -\sum_{i=1}^{m}\sum_{x=1}^{k(i)-1} nt_{i,x} * sb_{i,x} - \sum_{r=1}^{mscs} np_r * \ln \gamma_r$

und sb wird mittels m getrennten ($k_i$-1)- dimensionalen Newton-Raphson-Verfahren so ge-schätzt, dass cLL(sb) maximal wird. Beginnend mit sb = $\underline{0}$ und getrennt nach den m Zeilen wird sb = sb – verb wird solange iteriert, bis das Maximum aller absoluten Verbesserungen (vmax) kleiner als eps wird. Besteht der Verdacht, dass die Lösung nur ein lokales Maximum bildet, kann das Verfahren mehrmals mit zufälligen Startwerten wiederholt werden.
Die Verbesserungen für Item i lauten verb = $A^{-1}.sb'$, wobei $sb_{i,x}$ die ersten Ableitungen von cLL(sb) nach $sb_{i,x}$ ist. A ist die ($k_i$-1)- dimensionale Matrix der zweiten Ableitungen nach $sb_{i,x}$ und $sb_{iy}$ – diejenigen nach $sb_{i,x}$ und $sb_{jy}$ werden vernachlässigt. Nach jedem NR-Schritt wer-den die sb mittels Schwellenparametern normiert.
$\gamma_r$ ist symmetrische Grundfunktion (sgf) r- ter Ordnung und gleich der Summe aller Produkte exp[-$sb_{i,x(i)}$], bei denen die Summe aller $x_i$ gleich r ist. Die erste Ableitung von $\gamma_r$ nach $sb_{i,x}$ lautet $-sgf_{r,i,x}$ , wobei $sgf_{r,i,x}$ die Summe derjenigen Produkte ist, an denen $sb_{i,x}$ beteiligt ist. Man berechnet $sgf_{r,i,1}$ am einfachsten, indem man (vorübergehend) $\exp[-sb_{i,1}] = 0$ setzt, $\gamma^*$ berechnet und $\gamma_r^*$ von $\gamma_r$ abzieht. Die restlichen berechnen sich als $sgf_{r,i,1}*\exp[-sb_{i,x}+sb_{i,1}]$. Die zweiten Ableitungen von $\gamma_r$ nach $sb_{i,x}$ lauten $sgf_{r,i,x,x} = +sgf_{r,i,x}$ , diejenigen nach $sb_{i,x}$ und $sb_{i,y}$ sind $sgf_{r,i,x,y} = 0$, und diejenigen nach $sb_{i,x}$ und $sb_{j,x}$ (oder $sb_{j,y}$) werden vernachlässigt.

Die erste Ableitung von cLL(sb) nach $sb_{i,x}$ lautet $sb'_{i,x} = -nt_{i,x} + \sum_{r=1}^{mscs} np_r * sgf_{r,i,x} / \gamma_r$ .

Die 2. Ableitung nach $sb_{i,x}$ und $sb_{i,y}$ lauten $A_{x,y} = \sum_{r=1}^{mscs} np_r * (-sgf_{r,i,x,y} + sgf_{r,i,x} * sgf_{r,i,y}) / \gamma_r^2$ .

**Die symmetrischen Grundfunktionen**

Am einfachsten und genauesten ist die Berechnung, wenn mit Item 1 und $G_{1,.}$ begonnen und sukzessive ein weiteres Item dazu genommen wird, um $G_{i,.}$ zu berechnen. Die Länge von $G_{i,}$ ist die Summe der $k_j$-1 für j = 1 bis i. Am Ende ergeben sich die mscs Elemente von $\gamma = G_{m,}$ , wobei das letzte zur Kontrolle $\gamma_{mscs} = e^0 = 1$ sein muss. Da der Algorithmus verbal nur umständlich zu beschreiben ist, folgt ein Beispiel:

Die drei Items haben je 3, 2 und 4 Kategorien. $G_{1,}$ besteht aus den zwei Elementen 11 und 12, wobei z.B. $\exp[-sb_{2,1}]$ mit 12 abgekürzt ist.

Die Aufnahme von Item 2 ergibt $G_{2,1} = G_{1,1} + 21$, $G_{2,2} = G_{1,2} + G_{1,1}*21$ und $G_{2,3} = G_{1,2}*21$.

Mit Item 3 ergibt sich $G_{3,1} = G_{2,1} + 31$, $G_{3,2} = G_{2,2} + 32 + G_{2,1}*31$, $G_{3,3} = G_{2,3} + 33 + G_{2,2}*31 + G_{2,1}*32$, $G_{3,4} = G_{2,3}*31 + G_{2,2}*32 + G_{2,1}*33$, $G_{3,5} = G_{2,3}*32 + G_{2,2}*33$ und $G_{3,6} = G_{2,3}*33$.

Im Rechenprogramm wird $G_{i,r}$ nur einfach indiziert, und nach jeder Aufnahme $\gamma = G$ gesetzt.

### P14.8.9.3 Schätzung der gewichteten Personenparameter (fw)

Die Modellgleichung des partial credit model wird als $P(i,x_i|r) = Zl_{r,i} / Ne_{r,i}$ formuliert mit dem Zähler $Zl_{r,i} = \exp[x_i*fw_r - sb_{i,x(i)}]$ und dem Nenner $Ne_{r,i} = \sum_{y=0}^{k(i)-1} \exp[y * fw_r - sb_{i,y}]$.

Die <u>u</u>ngewichtete <u>L</u>ikelihood einer Person mit Scoresumme r lautet $uL(f_r) = \prod_{i=1}^{m} P(i, x_i \mid r)$.

Nach Warm gewichtet lautet sie $wL(fw_r) = \left( \prod_{i=1}^{m} P(i, x_i \mid r) \right) * \sqrt{-f_r''}$ ,

wobei $f_r''$ die 2. Ableitung der uL(f) nach $fw_r$ ist, d.h. nach Einsetzen des $fw_r$ in die uL.

Nach Konvergenz des Verfahrens ist $1/\sqrt{-f_r''}$ der geschätzte Standardfehler $se(fw_r)$. $f_r''$ wird im Verlauf des NR- Verfahrens $fw_r -= fw_r' / fw_r''$ als Nenner verwendet.

Nur für dichotome Items gilt $f_r'' = -\sum_{i=1}^{m} P(i, x_i \mid r) * \left(1 - P(i, x_i \mid r)\right)$.

Die gewichtete Loglikelihood lautet $wLL(f_r) = \sum_{i=1}^{m} \ln P(i, x_i \mid r) + \frac{1}{2} * \ln(-f_r'')$.

Die Summe wird zu $\sum_{i=1}^{m} \left( x_i * fw(r) - sb_{i,x(i)} - \ln Ne_{r,i} \right) = r * f_r - \sum_{i=1}^{m} \ln Ne_{r,i} + c$,

da die Summe der $x_i$ gleich r ist, und $sb_{i,x}$ nicht von r abhängt.

Das ergibt $wLL(fw_r) = r*fw_r - L_r + \ln(-f_r'')/2$, wobei $L_r = \sum_{i=1}^{m} \ln Ne_{r,i}$ .

Die Likelihood aller Personen mit Scoresumme r ist die $np_r$- te Potenz von $wL(fw_r)$. Die gewichtete Loglikelihood aller Beobachtungen lautet

$$wLL(\underline{fw}) = \sum_{r=0}^{mscs} np_r * \left( r * fw_r - L_r + \ln(-f_r'')/2 \right).$$

Die 1. Ableitung von wLL($fw_r$) nach $fw_r$ lautet $fw_r^{'} = r - L_r^{'} - f_r^{'''} / f_r^{''} / 2$,

wobei $L_r^{'} = \sum_{i=1}^{m} \frac{Ne_{r,i}^{'}}{Ne_{r,i}}$ und $Ne_{r,i}^{(\ell)} = \sum_{y=1}^{k(i)-1} y^{\ell} * \exp[y * fw_r - sb_{i,y}]$ gilt.

$L_r^{''} = \sum_{i=1}^{m} \left( \frac{Ne_{r,i}^{''}}{Ne_{r,i}} - \left( \frac{Ne_{r,i}^{'}}{Ne_{r,i}} \right)^2 \right) = -f_r^{''}$, wobei $1/\sqrt{-f_r^{''}}$ der geschätzte Standardfehler ist, und $-f_r^{''}$ als Schätzer von $fw_r^{''}$ verwendet wird (mit einem etwas kleineren Absolutbetrag). Seine Ableitung lautet $f_r^{'''} = \sum_{i=1}^{m} \left( Ne_{r,i}^{'''} - 3 * Ne_{r,i}^{''} * Ne_{r,i}^{'} / Ne_{r,i} + 2 * Ne_{r,i}^{'\,3} / Ne_{r,i}^{2} \right) / Ne_{r,i}$.

Der Iterationsschritt des NR- Verfahrens lautet somit fw– = fw'/fw''.
Als Startwerte eignen sich Nullen oder mscs+1 beliebige aufsteigend geordnete und summennormierte Zahlen. Das Verfahren konvergiert, wenn Max(|verb|) < ε. Der Mittelwert Mf weicht stets knapp von 0 ab. Berechnet man wLL mit den summennormierten fw-Mf, ergibt sich eine geringfügig kleinere Loglikelihood.

### P14.8.9.4 Standardfehler der Schwierigkeitsparameter sb

Setzt man sb und fw in die uLL ein, und bildet man die 2. Ableitung nach $sb_{i,x}$, ergibt sich

$$s_{i,x}^{''} = -\sum_{r=0}^{mscs} np_r * P(i,x \mid r) * (1 - P(i,x \mid r))$$

und der Standardfehler se($sb_{i,x}$) = $\sqrt{-1/s_{i,x}^{''}}$.

### P14.8.9.5 Andrichs Reliabilität:

Sie berechnet sich als ARel = 1 – mev/gVf

mit der marginal error variance = mev = $\sum_{r=0}^{mscs} np_r * sefw_r^2 / N$,

dem gewichteten Mittelwert der fw = gMf = $\sum_{r=0}^{mscs} np_r * fw_r / N$,

und der gewichteten Varianz der fw = gVf = $\sum_{r=0}^{mscs} np_r * (fw_r - gMf)^2 / N$.

### P14.8.9.6 Modell-Tests:

#### P14.8.9.6.1 Globale Modell-Tests:

**Anpassungstests (Pearsons Chi-Quadrat, Cressie-Read-Test)**

Ein allgemeines Abstandsmaß ist C(λ) = $\frac{2}{\lambda(\lambda+1)} * \sum_{j=1}^{kl} B_j * \left( \left( \frac{B_j}{E_j} \right)^{\lambda} - 1 \right)$, wobei kl die Zahl der Klassen ist, und $B_j$ und $E_j$ die beobachteten und erwarteten Häufigkeiten sind.

Für $\lambda = 1$ ergibt sich Pearsons $\chi^2$ : **PC** = $\sum_{j=1}^{amp} \frac{(pah_j - E(pat_j)^2}{E(pat_j)}$.

Für $\lambda = 2/3$ ergibt sich Cressie-Reads $\chi^2$ : **CR** = $1.8 * \sum_{j=1}^{amp} pah_j * \left( \left( \frac{pah_j}{E(pat_j)} \right)^{2/3} - 1 \right)$.

Beide Prüfgrößen sind asymptotisch $\chi^2$- verteilt mit df Freiheitsgraden. Deren Anzahl ist df = amp–2∗mscs, da mscs Parameter sb geschätzt worden sind, von denen einer wegen der Summennormierung entfällt. Weiters entfallen mscs+1 Freiheitsgrade wegen des Zusammenhangs $\sum_{scs(j)=r} E(pat_j)$ = $np_r$ = $\sum_{scs(j)=r} pah_j$.

Die erwarteten Patternhäufigkeiten berechnen sich als $E(pat_j) = np_r * \exp\left[ -\sum_{i=1}^{m} sb_{i,x(i)} \right] / \gamma_r$, wobei r $= \sum_{i=1}^{m} pat_{j,i}$ die Scoresumme ist. Da amp extrem groß werden kann, wird bei beiden Prüfgrößen nur über j = 1 bis n0 summiert. Da der Beitrag eines unbeobachteten Patterns zu PC bzw. CR gleich $E(pat_j)$ bzw. 0 ist, muss zu PC noch die Summe derjeniger Erwartungen addiert werden, bei denen $pah_j = 0$ gilt; sie lautet N – $\sum_{pah(j)>0} E(pat_j)$.

Sind mehrere Erwartungen kleiner als 5, sind die Prüfgrößen nicht $\chi^2$- verteilt, und p-values können nur mittels Bootstrappen geschätzt werden.

**Likelihood-Quotienten-Test**

Die log-ratio LR = $-2*(LL_{mod} - LL_{sat})$ ist asymptotisch $\chi^2$- verteilt mit df Freiheitsgraden.

$LL_{mod} = cLL(\underline{sb}) + \sum_{r=0}^{mscs} np_r * \ln(np_r / N)$ für $np_r > 0$.

$LL_{sat} = \sum_{j=1}^{n0} pah_j * \ln(pah_j / N)$, wobei n0 die Anzahl der besetzten Pattern ist.

### P14.8.9.6.2 Itemfit-Tests nach Wright & Masters

Out- und infit- Maße dienen zur Überprüfung der Modellannahme, dass die Trennschärfen der Items gleich sind. Sie verhalten sich unterschiedlich sensibel gegenüber bestimmten Modellverletzungen. Ihre t- Werte sollten standardnormalverteilte Prüfgrößen sein.

Der t- Wert des outfit- Maßes von Item i berechnet sich als $tout_i = \left( outfit_i^{1/3} - 1 \right) * c + 1/c$, wobei c $= 3/\sqrt{q2out_i / N^2 - 1/N}$. Im Multira-Programm von Carstensen/Rost wird vermutlich der Quotient 1/N nicht subtrahiert.

$outfit_i = \sum_{j=1}^{n0} pah_j * (pat_{j,i} - erw_i)^2 / var_i / N$. Da nicht die Trennschärfen der Items berechnet werden, sondern nur deren vom Modell geforderte Gleichheit untersucht wird, sind die m outfit-Maße derart zu korrigieren, dass ihr Mittelwert gleich 1 wird: $outfit_i$ + = 1 – Summe[$outfit_i$]/m. Ist $tout_i$

signifikant zu groß oder zu klein (z.B. größer 2 bzw. kleiner -2), schließt man mit ca. 95%- iger Sicherheit auf underfit bzw. overfit. Die Gleicheit der Trennschärfe ist dann nicht gegeben.

$q2out_i = \sum_{j=1}^{n0} pah_j * X4_i / var_i^2$ .

$X4_i = \sum_{x=0}^{k(i)-1} (x - erw_i)^4 * p_x$ , $var_i = \sum_{x=0}^{k(i)-1} (x - erw_i)^2 * p_x$ , $erw_i = \sum_{x=0}^{k(i)-1} x * p_x$ ,

wobei $p_x$ die Modellwahrscheinlichkeit $P(i,x_i|r)$ ist, und r die Scoresumme von $pat_j$ .

Die t- Werte der infit- Maße berechnen sich als $tin_i = \left(infit_i^{1/3} - 1\right) * c + 1/c$ ,

wobei $c = 3/\sqrt{q2in_i}$ und $q2in_i = \sum_{j=1}^{n0} pah_j * (X4_i - var_i^2) / \left( \sum_{j=1}^{n0} pah_j * var_i \right)^2$ .

Die m Maße $infit_i = \sum_{j=1}^{n0} pah_j * (pat_{j,i} - erw_i)^2 / \sum_{j=1}^{n0} pah_j * var_i$

sind noch wie vorhin zu korrigieren: $infit_i$ + = 1 – Summe[$infit_i$]/m.

Man beachte: Die t- Werte der outfit- und infit-Maße sind nicht standardnormal verteilt, wie folgende drei jederzeit wiederholbare Versuchsanordnungen mit dichotomen Items zeigen:
a) 1000 Mal wird wiederholt: Simulation von m normierten Schwierigkeitsparametern SB ~ U(-2,2), Simulation von m+1 summennormierten und sortierten Fähigkeitsparametern FW ~ N(0,2), $NP_i$ =50, Simulation der Daten, Schätzung der Modellparameter, Berechnung der tout- und tin-Werte und Abspeicherung in 2 Matrizen der Dimensionen 1000∗m.
b) Wie Methode a), jedoch werden die Parameter SB und FW als arithmetische Folgen zwischen -2 und 2 bzw. -2.5 und 2.5 festgehalten, d.h. nicht mehr simuliert.
Bei Betrachtung der 2∗m Spalten, die Stichproben aus N(0,1)- Verteilungen sein sollten, stellt sich (von zufallsbedingten geringen Abweichungen abgesehen) Folgendes heraus:

a) m = 5: $tout_i$ ~ N(0.1, $0.35^2$), $tin_i$ ~ N(0.1, $0.75^2$)
m = 10: $tout_i$ ~ N(0.05, $0.5^2$), $tin_i$ ~ N(0.1, $0.85^2$)

b) Die Mittelwerte und Varianzen hängen von den Positionen der SB ab. Beide sind bei den Items 1 und m minimal und steigen zum mittelschweren Item hin an. Dies ist in den folgenden 4 Zeilen mittels „min→max“ beschrieben:
m = 5: $tout_i$ ~ N( 0.10→0.20, $0.30^2$→$0.40^2$ )
$tin_i$ ~ N( -0.25→0.35, $0.40^2$→$0.65^2$ )
m =10: $tout_i$ ~ N( 0.00→0.10, $0.50^2$→$0.60^2$ )
$tin_i$ ~ N( -0.20→0.10, $0.60^2$→$0.85^2$ )

c) Die Parameter SB werden als arithmetische Folgen zwischen -2 und 2 festgelegt, und es werden 30000 bzw. 55000 exakte Daten erzeugt – d.h. die Patternhäufigkeiten (pah) stimmen bis auf Rundungsfehler mit ihren Erwartungen überein:
m = 5: $outfit_i$ = 0.97→1.03, $infit_i$ = 0.97→1.02
$tout_i$ = -0.93→1.94, $tin_i$ = -2.77→3.38
m = 10: $outfit_i$ = 0.97→1.02, $infit_i$ = 0.98→1.01
$tout_i$ = -1.27→1.80, $tin_i$ = -2.10→1.54

### P14.8.9.6.3 Der Q-Index (nach Rost & von Davier, Appl. Psy. M. 1994, 171-182)

Die Q-Indizes sind Maße dafür, wie weit die beobachteten von den optimalen Pattern abweichen und sollten modellgemäß gleich groß sein. Je Item werden 3 Vektoren der Länge N erzeugt: im ersten stehen die Scoresummen r, im zweiten die Beobachtungen x = $pat_i$ . Jetzt werden die Scoresummen aufsteigend sortiert, und mit ihnen die x mit demselben r.
Im dritten Vektor stehen die Werte g und entsprechen den Guttmann-Pattern: für sie gilt $g_n = 0$ für $1 \le n \le nt_{i,0}$ , $g_n = 1$ für $nt_{i,0} < n \le nt_{i,0}$ usw. bis $g_n = k_i\text{-}1$ für die letzten $nt[i,k_i\text{-}1]$ Werte. Die Anti-Guttman-Werte ergeben sich als $ag_n = g_{N+1-n}$ .und werden zur Normierung gebraucht.
Der Q-Index des Items i berechnet sich als $Q_i = \sum_{n=1}^{N}(x_n - g_n) * fw_{r[n]+1} / \sum_{n=1}^{N}(x_n - ag_n) * fw_{r[n]+1}$ .
Im oben angeführten Artikel werden außerdem noch $z(Q_i)$ hergeleitet, die standardnormalver-teilte Prüfgrößen sein sollten. Die im Aufsatz von Rost und Davier angegebenen Formeln konnten wir jedoch nicht programmieren bzw. ihre Programmierung hat zu unbrauchbaren Ergebnissen geführt. Vermutlich sind die Formeln unvollständig.

### P14.8.9.6.4 Personenhomogenität (Andersen-Test, Wald-Test)

Nach Teilung der Personen in die Gruppen r ≤ split und r > split werden in beiden Gruppen die Itemparamer sb1 und sb2 geschätzt und die loglikelihoods cLL1 und cLL2 berechnet. Die Andersen-Prüfgröße -2(cLL-cLL1-cLL2) ist asymptotisch $\chi^2$- verteilt mit mscs-1 Freiheits-graden. Um festzustellen, welche Items eine eventuelle Inhomogenität verursachen, werden die Wald-Statistiken $z_{i,x} = (sb1_{i,x} - sb2_{i,x}) / \sqrt{se(sb1)_{i,x}^2 + se(sb2)_{i,x}^2}$ berechnet. Sie sind nur bei dichotomen Items N(0,1)- verteilt. Weder die Summe $\sum_{i=1}^{m} z_{i,1}^2$ bei dichotomen noch $\sum_{x=1}^{k(i)-1} z_{i,x}^2$ bei polytomen Items ist $\chi^2$- verteilt.

### P14.8.9.6.5 Itemhomogenität (Martin-Löf-Test)

Für den Martin-Löf-Test werden die Items entsprechend zweier Indexmengen in Gruppen ge-teilt. Die Maximalscores in beiden Gruppen lauten msc1 und msc2. Es werden die Itempara-meter sb1 und sb2 geschätzt und die loglikelihoods $cLL_1$ und $cLL_2$ berechnet. Die Prüfgröße X2 = -2∗($LL_{mod}$ – $LL_{12}$) ist asymptotisch $\chi^2$- verteilt mit msc1∗msc2-1 Freiheitsgraden.

$LL_{mod}$ = cLL(sb) + $\sum_{r=0}^{mscs} np_r * \ln(np_r / N)$.

$LL_{12}$ = cLL(sb1) + cLL2(sb2) + $\sum_{r1=0}^{msc1}\sum_{r2=0}^{msc2} n_{r1,r2} * \ln(n_{r1,r2} / N)$,

wobei $n_{r1,r2}$ die Häufigkeit der Scoresummen r1 und r2 in den beiden Gruppen ist.

### P14.8.9.7 Bootstrappen

Die Vorgangsweise, nach welcher der p-value der globalen Prüfgrößen (1)Pearsons Chi-Quadrat PC, (2)Cressie-Read CR, (3)log-ratio LR parametrisch bootsgetrappt wird, ist folgende:

1. Die zu prüfende Statistik wird als X0 festgehalten, die Parameter sb,fw,np als SB,FW,NP.
2. Es wird eine Zahl B gewählt und B Mal werden neue Daten mittels SB, FW und NP simuliert – im Gegensatz zum originalen Verfahren, wo B Mal mit Zurücklegen aus den Daten gezogen wird. In jeder der B Stichproben werden
2a) sb geschätzt,
2b) die Statistik XB berechnet,
2c) und wird mitgezählt, wie oft XB > X0 ist; die Anzahl ist $0 \leq nB \leq B$.
3. Die relative Häufigkeit pv = nB/B ist nun die bootsgetrappte Überschreitungswahrscheinlichkeit, dass der wahre Wert X0 unter $H_0$ größer als das beobachtete X0 ist. Ist p der wahre Wert, dann ist pv Bin(B,p)- verteilt. Ist B∗pv∗(1-pv) hinreichend groß, dann ist pv annähernd N( pv, B∗pv∗(1-pv) )- verteilt, womit ein Konfidenzintervall angegeben werden kann.

Die p-values sind die Wahrscheinlichkeiten von Fehlern 1. Art, d.h. $H_0$ irrtümlich abzulehnen. Eine notwendige Bedingung, dass sie richtig geschätzt worden sind, ist die, dass sie der Stich-probe einer U(0,1)- Verteilung entsprechen müssen (Die Transformation $Y = F_x(X)$ ist gleichverteilt). Dies kann man wie folgt nachprüfen:
Man nimmt m und k an, sowie die Verteilungen von SB0,FW0,NP0, und wiederholt folgende 4 Vorgänge wh Mal.

1. Simulation von SB0,FW0,NP0
2. Schätzung der sb,fw und Festhalten der SB,FW,
3. Berechnung der Prüfgröße und Festhalten als X0,
4. B Mal Bootstrappen und Speichern des p-values.

Das ergibt wh Beobachtungen pv, die mittels Histogramm graphisch und mittels Verteilungstest statistisch geprüft werden können.

### P14.8.9.8 Simulation von „Rasch-konformen“ Daten.

Für r = 0,…,mscs werden jeweils $NP_r$ Pattern gemäß der Modellgleichung P(i,x|r) simuliert. Es wird eine Matrix P für $1 \leq i \leq m$ und $0 \leq x \leq k_i-1$ berechnet und zeilenweise kumuliert. Jetzt wird je Item eine standardgleichverteilte Zufallszahl (zuf) erzeugt und die Stufe $pat_i$ bestimmt, für die $P_{i,x-1} <$ zuf $< P_{i,x}$ gilt. Daraus werden die dezN = pat.bas gebildet.
Ist man an Pattern interessiert, deren Häufigkeiten exakt den theoretischen entsprechen, werden alle (d.h. n0 = amp) pah berechnet, indem das Produkt der Zeile $P_{i,x}$ mit $NP_r$ multipliziert wird. Schätzt man mit diesen Daten die Parameter, ergibt sich sb = SB und fw ≠ FW. Die fw sind stets dieselben, egal wie FW und NP gewählt worden sind – sogar wenn FW = $\underline{0}$ und alle NP = 0 mit Ausnahme irgendeines $NP_r$ = N. Die globalen Testgrößen und die des Martin-Löf-Test sind exakt Null, die Itemfitmaße und Q-Indizes jedoch nicht – die t-Werte werden bei großen Stichproben signifikant.

Kurt Holm

## Exkurs: Zentrale Konzepte des Rasch-Modells

Die im Text an diversen Stellen behandelten zentralen Konzepte des Rasch-Modells sollen nun nochmals in kompakter Form vorgetragen werden. Dabei sollen eher pragmatische Definitionen verwendet werden.

**0. Logistische Itemfunktion**

Dies ist die Grundannahme des Rasch-Modells. Das Rasch-Modell unterlegt den dichotomen und polytomen Items eine vereinfachte 1-parametrische logistische Funktion. Siehe Abschnitt P14.8.3.1, Formel 0 und Abschnitt P14.8.3.2, Formel 1. Im Grunde genommen gehen alle nachfolgend aufgeführten Konzepte zwingend aus dieser Grundannahme hervor.

**1. Eindimensionalität.**

Die Items müssen bei den Untersuchungspersonen die gleiche Dimension und nur diese eine messen. Ob diese Bedingung erfüllt wird, kann durch eine Faktorenanalyse und ungefähr durch den Vergleich des Rasch-Modells mit dem multidimensionalen Rasch-Modell überprüft

werden. Bei polytomen Items setzt die Faktorenanalyse allerdings quantitative Messwerte und nicht ordinale wie beim Rasch-Modell voraus. Beim multidimensionalen Rasch-Modell muss der Forscher über die Dimensionszugehörigkeit der Items zuvor eine Entscheidung treffen. Das multi-dimensionale Rasch-Modell überprüft dann, ob seine Entscheidung akzeptabel war.

**2. gleiche Trennschärfe.**
Die logistischen "Kurven der Antwort-Wahrscheinlichkeit" der dichotomen Items müssen (annähernd) gleich steil verlaufen. In Gleichung L2 in Abschnitt P14.8.3.1 ist dies der Parameter c, von dem verlangt wird, dass er für alle Items (annähernd) gleich groß ist. Siehe dazu Abschnitt P14.8.7.10 den Punkt "4. Eingabefeld: Trennschärfe (Steigung) der Items". Diese Bedingung kann durch die Itemfit-Maße, insbesondere durch "Outfit" und Infit" überprüft werden.

*"angemessene" Steilheit der Itemkurve.*
Ist die Steigung sehr flach oder sehr steil dann sollten die Ergebnisse des Rasch-Modells nicht mehr interpretiert werden. Die Angemessenheit der Steigung wird ungefähr durch den Q-Index nach Rost und von Davier ermittelt.

**3. Separierbarkeit von Schwierigkeitsparameter des Items und Fähigkeitsparameter der Person**.
Schwierigkeit ist ausschliesslich eine Eigenschaft der Items und Fähigkeit eine Eigenschaft der Personen. Schwierigkeitsparameter und Fähigkeitsparameter werden im Rasch-Modell auf derselben Skala abgetragen. Grafisch betrachtet: Sie werden gemeinsam auf der x-Achse des logistischen Koordinatensystem abgetragen. Trotzdem können sie unabhängig voneinander geschätzt werden. Sind Item- und Personenparameter separierbar, dann ist die Messung "spezifisch objektiv". Eine sehr technische Definition des Konzepts der Separierbarkeit gibt Sixtl in Befragung 4, S. 28 ff.

**4. spezifische Objektivität**
Wir definieren diesen zentralen Begriff des Rasch-Modells pragmatisch durch folgende 2 Sätze:

(1) Die Schwierigkeitsparameter der Items sind unabhängig davon welchen Personen die Items zur Beantwortung vorgelegt wurden

Das kann durch den Andersen-Test für Personenhomogenität gemessen werden. Für 2 Gruppen von Personen müssen die Schwierigkeitswerte derselben Items gleich sein (bzw. sie dürfen nicht signifikant verschieden sein)

(2) Die Fähigkeitsparameter der Personen sind unabhängig davon welche Items (natürlich zur Messung derselben Dimension) den Personen zur Beantwortung vorgelegt wurden.

Das kann durch den Martin-Löf-Test für Itemhomogenität gemessen werden. 2 verschiedene Mengen von Items (natürlich zu derselben Messdimension) müssen für dieselben Personen die gleichen Fähigkeitswerte erbringen (bzw. sie dürfen nicht signifikant verschieden sein)

Die spezifische Objektivität ist ein generelles Prinzip des Messens und nicht nur ein Prinzip des Rasch-Modells.

Beispiel:
1. Die Temperatur in einem Glas Wasser lässt sich mit einem Quecksilber-Thermometer messen, aber auch mit einem Alkohol-Thermometer oder mit einem Bimetall oder Halbleiter-Temperatursensor. Das Objekt hat eine Temperatur in Grad Celsius (oder Fahrenheit oder sonstiger Maßeinheit), die von verschiedenen Messinstrumenten objektiv gemessen wird.
2. Ein Thermometer misst die Temperatur eines Glases, das mit Wasser gefüllt ist, aber auch die eines Glases, das mit Rotwein gefüllt ist. Der Thermometer weist den beiden Objekten in derselben Dimension einen vergleichbaren Messwert zu.

Auf das sozialwissenschaftliche Messen angewendet bedeutet dies:
Die Items einer Rasch-Skala müssen auf eine Personenstichprobe A so anwendbar sein wie auf eine Personenstichprobe B. Konkreter: Die einzelnen Items einer Rasch-Skala müssen denselben Schwierigkeitswert für Personenstichprobe A und für Personenstichprobe B besitzen. Weniger rigoros formuliert: Die einzelnen Items einer Rasch-Skala müssen mindestens dieselbe Rangfolge der Schwierigkeitswerte für Personenstichprobe A wie für Personenstichprobe B besitzen.

*Spezifische Objektivität und Pisa-Studie.*
Betrachten wir als Beispiel den Mathematik-Test aus der Pisa-Studie. Siehe hierzu auf
`www.almo-statistik.de`
die Seite "Pisa-Studie 2012" und hier "Pisa2012_Mathe_A_D_Shanghai". Nachdem dieser Ordner herunter geladen ist, sollte er in den Almo-Ordner kopiert werden.

a. Zunächst wird unterstellt, dass "mathematische Fähigkeit" eine zwar latente (also nicht direkt zugängliche) Dimension ist - die existiert und damit beim Schüler aus Österreich dieselbe ist wie beim Schüler aus Shanghai.
b. Die Grundvoraussetzung ist, dass die "Kurve der Antwort-Wahrscheinlichkeit" für alle Untersuchungspersonen dieselbe Form besitzt, damit diese vergleichbar gemessen werden können.
c. Das Rasch-Modell beansprucht dann, mit der logistischen Form diese Dimension vergleichbar zu messen
d. Um diese Leistung erbringen zu können, sind folgende 2 Annahmen notwendig:
(1) die eingesetzten Test-Items besitzen dieselbe Schwierigkeit z.B. für den Schüler aus Österreich wie für den Schüler aus Shanghai und (2) die Trennschärfe der Items also die Steigungsparameter der Funktionen ist für die beiden Schülergruppen bei jedem einzelnen Item derselbe

Wir haben ein Rasch-Modell der Mathe-Items aus Pisa 2012 des Booklet 10 für Austria und Shanghai gerechnet. Dabei wurden nur die Daten der tatsächlich befragten 355 österreichischen Schüler und der 300 Shanghaier Schüler verwendet. Es wurde ein Andersen-Test für die Gruppierungsvariable "Land" gerechnet. Die österreichischen und Shanghaier Schüler erwiesen sich dabei als nicht "Personen-homogen". Die Schwierigkeitswerte der Items sind für die beiden Gruppen eklatant verschieden. Die Folgerung daraus: Es besteht keine "spezifische Objektivität" des Rasch-Messverfahrens hinsichtlich dieser Daten.

**5. lokale stochastische Unabhängigkeit**

Innerhalb einer Gruppe von Personen mit gleicher Fähigkeit korrelieren die Items nicht signifikant. Egal welchen Schwierigkeitswert ein Item i besitzt, es erbringt für dieselbe Person den gleichen Fähigkeitswert wie das Item j, das einen anderen Schwierigkeitswert besitzt.

**6. Suffizienz des Gesamtpunkts.**

Mehrere Personen mit dem gleichen Gesamtpunkt erhalten den gleichen Fähigkeitswert - unabhängig davon welche Items bzw. Itemkategorien sie positiv beantwortet haben.

Konkret bedeutet dies, um ein Rasch-Modell rechnen zu können, werden folgende empirische Daten benötigt. Wir verwenden die Daten aus der Programm-Maske Prog14m4

(1) Die Zahl der Personen je Gesamtpunkt
Das ist ein Vektor, der z.B. folgendermaßen aussehen kann

```
Zahl der Personen
je Gesamtpunkt (in Klammern)
      7    (0)
     27    (1)
     55     .
     84     .
    124     .
    113
    130
    111
     58     .
     54     .
     31     .
     18   (11)
      3   (12)
```

(2) Die Zahl der Personen je Auspraegung je Item
Das ist eine Matrix, die z.B. folgendermaßen aussehen kann

| | Ausprägung 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Item 1 | 171 | 644 | - | - |
| Item 2 | 293 | 522 | - | - |
| Item 3 | 249 | 410 | 156 | - |
| Item 4 | 281 | 401 | 133 | - |
| Item 5 | 107 | 472 | 164 | 72 |
| Item 6 | 187 | 388 | 178 | 62 |

```
Item 1 und 2 besitzen die Ausprägungen 0 und 1
Item 3 und 4 besitzen die Ausprägungen 0, 1 und 2
Item 5 und 6 besitzen die Ausprägungen 0, 1, 2 und 3
```

Mehr an Daten werden nicht benötigt. Die Personenzahl je Gesamtpunkt und je Ausprägung sind "suffizient" um die Fähigkeitswerte der Individuen zu ermitteln. Es werden also nicht die konkreten Antwortmuster der Personen gebraucht. Sollen jedoch Modelltests oder Itemfit-Maße

ermittelt werden, dann müssen sehr wohl die Antwortmuster und deren Besetzungshäufigkeiten bekannt sein.

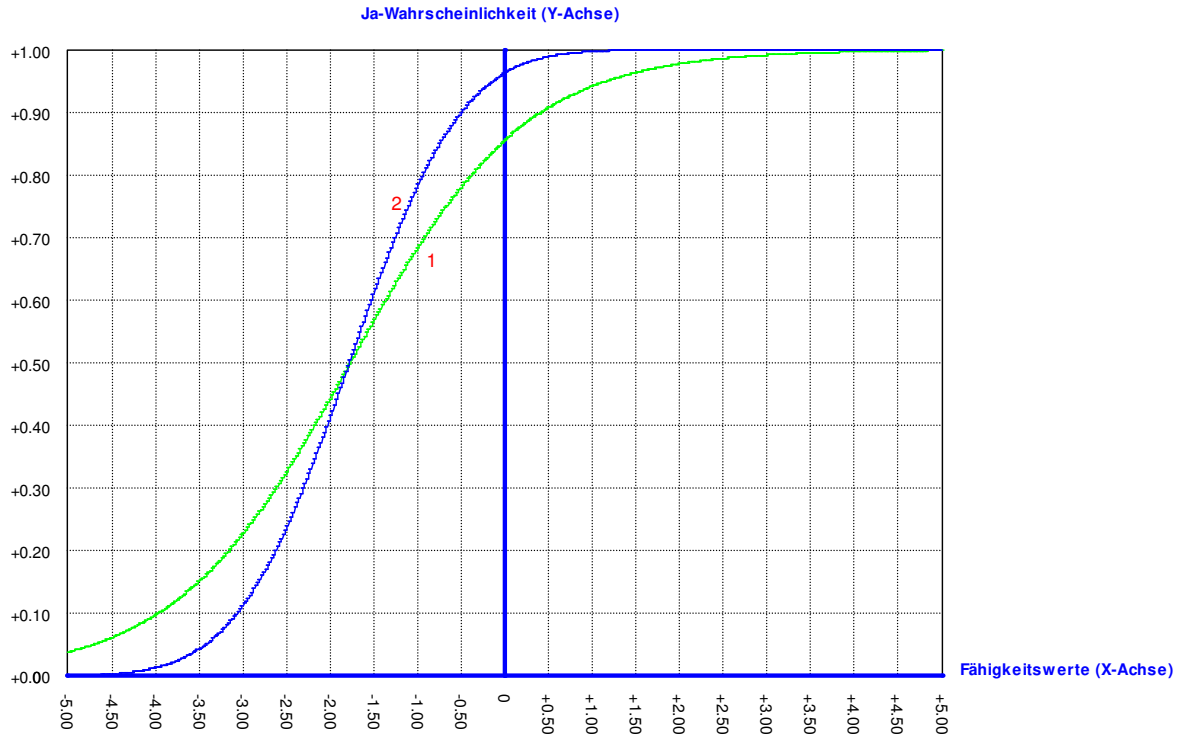
G.H. Fischer (1995) fasst die verschiedenen Annahmen in 5 kompakt formulierten Bedingungen für das Rasch-Modell zusammen. Siehe dazu

Fischer, G.H. und Molenaar, I.W. 1995. Rasch models - Foundations, recent developments, and applications. New York: Springer

Kurt Holm

**Exkurs: Vergleich von Rasch-Modell mit "Latent Structure Analysis" (LSA)**

Die "Latent Structure Analysis" besteht aus 5 Modellen. Uns interessiert hier das "Ogiven-Modell". Es verwendet anstelle der 1-parametrischen logistischen Funktion die 2-parametrische Ogiven-Funktion. Die beiden Funktionen sind grafisch dargestellt fast deckungsgleich.




Die grüne Kurve 1 ist die logistische Funktion des dichotomen Rasch-Modells mit dem Schwierigkeitswert -1.78. Der Trennschärfekoeffizient (die Steigung) ist im 1-parametrischen Rasch-Modell implizit =1.0

Die blaue Kurve 2 ist die Ogiven-Funktion der latent structure analysis mit demselben Schwierigkeitswert -1.78 und derselben Steigung =1.0.

In Almo erreicht man das Ogiven-Modell der "Latent Structure Analysis" durch Klick auf den Knopf "Verfahren", dann "Skalierungsverfahren" und Prog15ma. In der Almo-Ausgabe zu Prog15ma wird der Schwierigkeitsparameter mit m(i) bezeichnet und die Steigung mit a(i). Das Ogiven-Verfahren ist nur auf dichotome Items anwendbar.

Das von Hermann Denz für Almo programmierte Verfahren ist ausführlich beschrieben in seinem Buch "Analyse latenter Strukturen" (1982). Eine kurze Version ist im Almo-Handbuch "Sozialwissenschaftliche Skalierungsverfahren" vorhanden, sowie als selbständiges Dokument "Latent Structure Analysis.PDF". Beide, das Handbuch und das PDF-Dokument sind im Internet bei

www.almo-statistik.de

herunterladbar. Siehe auch nachfolgende Literaturliste. Das PDF-Dokument empfehlen wir zuerst zu lesen.

Wir werden im folgenden die Bezeichnugn "Latent Structure Analysis" durch LSA abkürzen

Die Ogive entsteht aus der Integration der Normalverteilungsfunktion
Die Formel für die Ogive der LSA lautet:

```
(1)  Y = Integral( c/2.507 * e**(- (c*(X-S)) * (c*(X-S)) / 2))
```

In gewohnter Schreibweise

$$(1a) \qquad Y_{ig} = \int_{-\infty}^{X} \frac{c_i}{2.507} * e^{-\left(\frac{c_i*(X_g-s_i)\ *\ c_i*(X_g-s_i)}{2}\right)}$$

```
   Die Integration erfolgt von minus unendlich bis X(ig).
   2.507 ist Wurzel aus 2*Pi.

   Wie beim Rasch-Modell ist die Obergrenze, der sich die Kurve
   asymptotisch nähert bei y=1.0, die Untergrenze bei y=0.

** bedeutet potenzieren

i  Index für Item i

g  Gesamtpunkt

Y  ist die Wahrscheinlichkeit, dass die Person mit Gesamtpunkt g das
   Item i positiv beantwortet

c  bestimmt die Steilheit des Items i. Je grösser c ist umso steiler.
   c kann interpretiert werden, als die Diskriminationsfähigkeit
   der Variablen, d.h. als Trennschärfe-Koeffizient.
   c ist der Kehrwert der Standardabweichung st, also

   c=1/st

st Standardabweichung der Werte von Item i

X  ist der Fähigkeitsparameter der Person mit Gesamtpunktwert g

S  ist der Schwierigkeitsparameter des Items i.

   Der Wendepunkt der Ogive liegt wie beim Rasch-Modell bei y=0.5.
   und über dem Schwierigkeitswert S des Items; bzw. der X-Wert, über
dem
   Y 0.5 groß ist, wird als Schwierigkeitswert des Items definiert.
```

Werden die X-Werte der Personen in standardisierten z-Werten ausgedrückt (d.h. st wird zur Maßeinheit), dann ergibt sich obige Gleichung (1) in folgender Form

```
(2)  Y=Integral( 1/2.507 * e**(z*z/2) )

Die Integration erfolgt von minus unendlich bis z.
2.507 ist Wurzel aus 2*Pi.
```

In gewohnter Schreibweise

$$(2a) \qquad Y_{ig} = \int_{-\infty}^{z_{ig}} \frac{1}{2.507} * e^{-(\frac{z_{ig} * z_{ig}}{2})}$$

```
Der z-Wert entsteht aus
```

$$z_{ig} = c_i * (X_g - s_i) = \frac{X_g - s_i}{st_i}$$

```
wobei X =Fähigkeitswert der Person mit Gesamtpunktwert g
      s =Schwierigkeitswert des Items i
      st=Standardabweichung des Items i;
```

Das ist die Formel die Denz (1982, S.76) verwendet

Der entscheidende Vorteil der LSA gegenüber dem Rasch-Modell besteht darin, dass die LSA explizit die Trennschärfe der Items in das Modell aufnimmt und unterschiedliche Trennschärfen erlaubt. Wie Denz ausführt, erkauft die LSA diesen Vorteil damit, dass sie nicht "spezifisch objektiv" mißt, so wie dies für das Rasch-Modell behauptet wird. Schwierigkeitsparameter der Items und Fähigkeitsparameter der Personen sind nicht separierbar. Die Schwierigkeitsparameter der LSA sind abhängig von der gewählten Stichprobe. Siehe dazu Sixtl (1976, S. 28 ff).

Ein weiteres ernsthaftes Problem der 2-parametrischen Modelle ist es, dass sich die Kurven im Bereich zwischen minimalen und maximalen Fähigkeitswert schneiden können (aber nicht müssen). Vor dem Schnittpunkt zweier Kurven ist das Item mit dem größeren Schwierigkeitswert (wie dies sein muss) schwieriger zu beantworten. Seine Kurve verläuft korrekt unterhalb der des leichteren Items. Nach dem Schnittpunkt ist dann das Item mit dem niedrigeren Schwierigkeitswert schwerer zu beantworten. Das ist inkonsistent. In der "Mokken-Analyse" wird diese Inkonsistenz durch den H-Wert ausgedrückt. Wir werden das noch zeigen.

Es muss hier noch auf das "Birnbaum-Modell" hingewiesen werden, das eine 2-parametrische logistische Funktion verwendet und ebenfalls die Trennschärfe explizit beachtet. In den Abschnitten P14.8.3.1 und P14.8.7.10 haben wir auf dieses Verfahren hingewiesen. Es ist in Almo nicht enthalten.

Die LSA wurde schon sehr früh durch die Beiträge einer Vielzahl von Autoren entwickelt. Systematische Darstellungen sind z.B. enthalten in

W.S. Torgerson: Theory and Method of Scaling, Wiley, 1960, besonders Kap. 13
P.F. Lazarsfeld & N.W. Henry: Latent Structure Analysis, Boston, 1968

Eine (gewichtete) Kleinste-Quadrate-Schätzung für das Ogiven-Modell wurde insbesondere entwickelt von

L.R. Tucker: A level of proficiency scale for a unidimensional skill, Amer. Psychologist, 7,408

Dieses Tucker-Verfahren ist im Almo-Ogiven-Modell der LSA eingesetzt. Eine umfangreiche Literaturliste ist bei Denz (1982) zu finden.

Eine Berechnung des Ogiven-Modells "von Hand" war zu jener Zeit sehr aufwendig bis unmöglich. Es wurde deswegen mit Erleicherung aufgenommen, als Rasch im Jahr 1960 sein 1-parametrisches logistisches Modell vorstellte, das weniger aufwendig (aber immer noch mühsam genug) "von Hand" zu rechnen war. Zu diesem Zeitpunkt standen dann auch schon mehr Groß-Computer zur Verfügung. Der PC war noch weit entfernt.

In Almo ist die LSA enthalten. Der leider zu früh verstorbene Hermann Denz hat sie mit allen Submodellen programmiert. In Almo sind die Maskenprogramme zur LSA erreichbar durch Klick auf den Knopf

```
"Verfahren / Skalerungsverfahren / Prog15m1" (alle Modelle)
oder "...Prog15ma" (das Ogiven-Modell, das hier interessiert)
```

Wir haben nun aus den Mathe-Daten der Pisa-Studie 2009 7 Items herausgegriffen und eine LSA und ein Rasch-Modell gerechnet.Siehe die Beispiel-Programme "Rasch_PisaMathe_7Items.Alm" und "LatStruc_PisaMathe_7Items.Alm". Erreichbar über das Menü "Almo", dann "Liste aller Almo-Programme".

Die Programme liefern folgende Schwierigkeits- und Fähigkeitsparameter:

Schwierigkeits-
parameter

| Variable | Rasch-Modell | Ogiven-Modell |
|---|---|---|
| 1 M033Q01.MATH.P | -1.620 | -1.389 |
| 3 M155Q01.MATH.P | -1.135 | -0.735 |
| 4 M155Q04T.MATH. | -0.294 | -0.169 |
| 5 M192Q01T.MATH. | 0.366 | 0.157 |
| 6 M406Q01.MATH.P | 0.681 | 0.353 |
| 7 M406Q02.MATH.P | 1.724 | 0.904 |
| 9 M411Q02.MATH.P | 0.279 | 0.331 |

Faehigkeitswerte

| Gesamtpunktwert | Rasch-Modell | Ogiven-Modell |
|---|---|---|
| 0 | -3.2610 | -1.8340 |
| 1 | -1.8611 | -1.2898 |
| 2 | -1.0161 | -0.7457 |

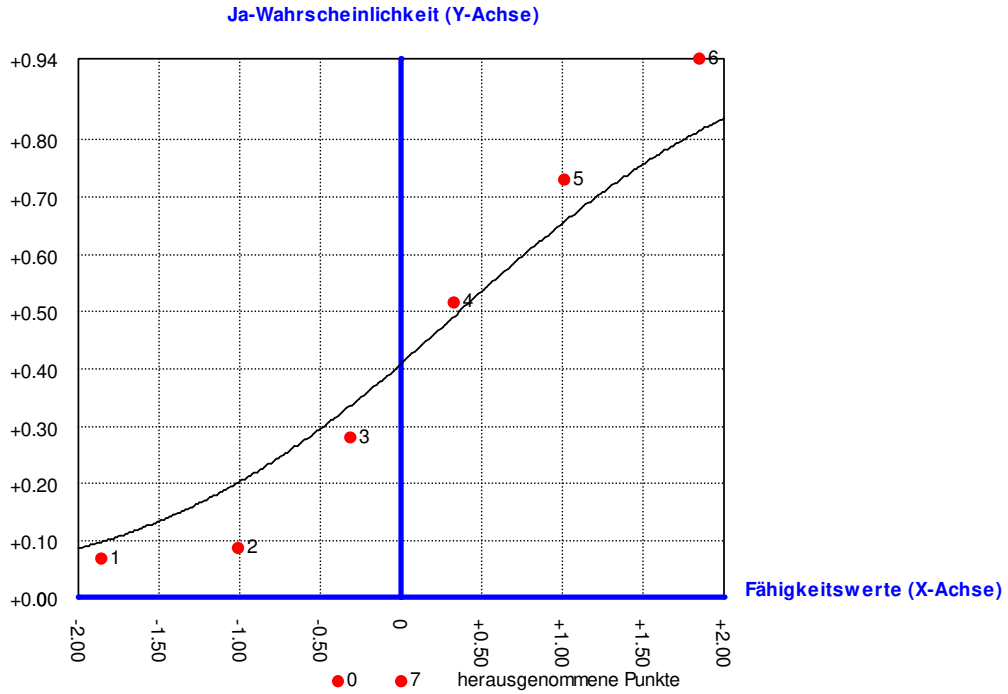
| | | |
|---|---|---|
| 3 | -0.3181 | -0.2349 |
| 4 | 0.3338 | 0.2050 |
| 5 | 1.0161 | 0.6516 |
| 6 | 1.8477 | 1.2562 |
| 7 | 3.2563 | 1.8609 |

Sowohl Schwierigkeits- als auch Fähigkeitsparameter sind verschieden skaliert, so dass sie nicht direkt wertmäßig verglichen werden dürfen. Die Werte aus dem Rasch-Modell sind sozusagen stärker "gespreizt".

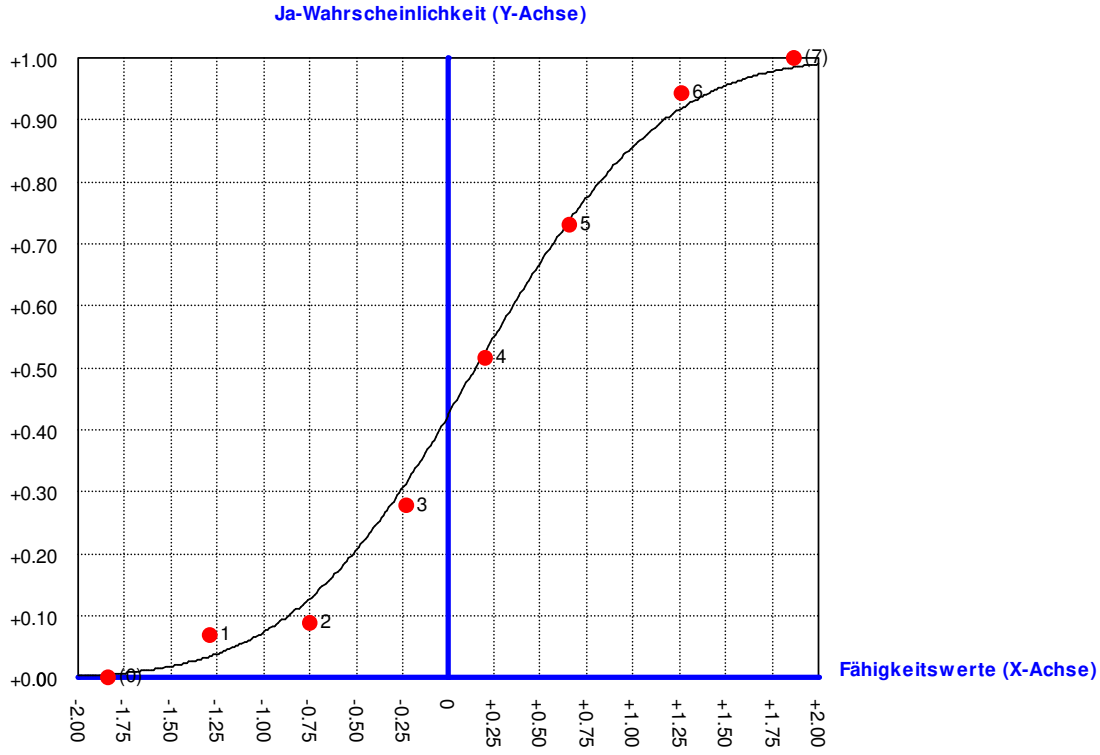
Die Ergebnisse grafisch dargestellt: Wir zeigen nur die Kurven für V5.
Beim Vergleich des Kurvenverlaufs der beiden Verfahren ist zu beachten, dass Almo beim Rasch-Verfahren die Kurve oben und unten abschneidet, da die Extremwerte 0 und 7 nicht miteinbezogen werden. Bei der LSA werden jedoch extrapolierte Schätzwerte für 0 und 7 eingesetzt. Ersichtlich ist jedoch, dass sich die Ogive dichter an die empirischen Punkte anschmiegt als die logistische Kurve. Das muss aber nicht immer so sein.

Die beiden Kurven sind folgende:

```
Logistische Funktion Y = 1/(1+e**(0.366-X)) für Variable V5 aaM192Q01T.MATH.
rote Punkte=empirische Ja-Wahrscheinlichkeit bei einem Gesamtpunktwert von x
Die Ges.punktwerte stehen bei den Punkten. Ges.punktwerte in Klammern geschätzt
```

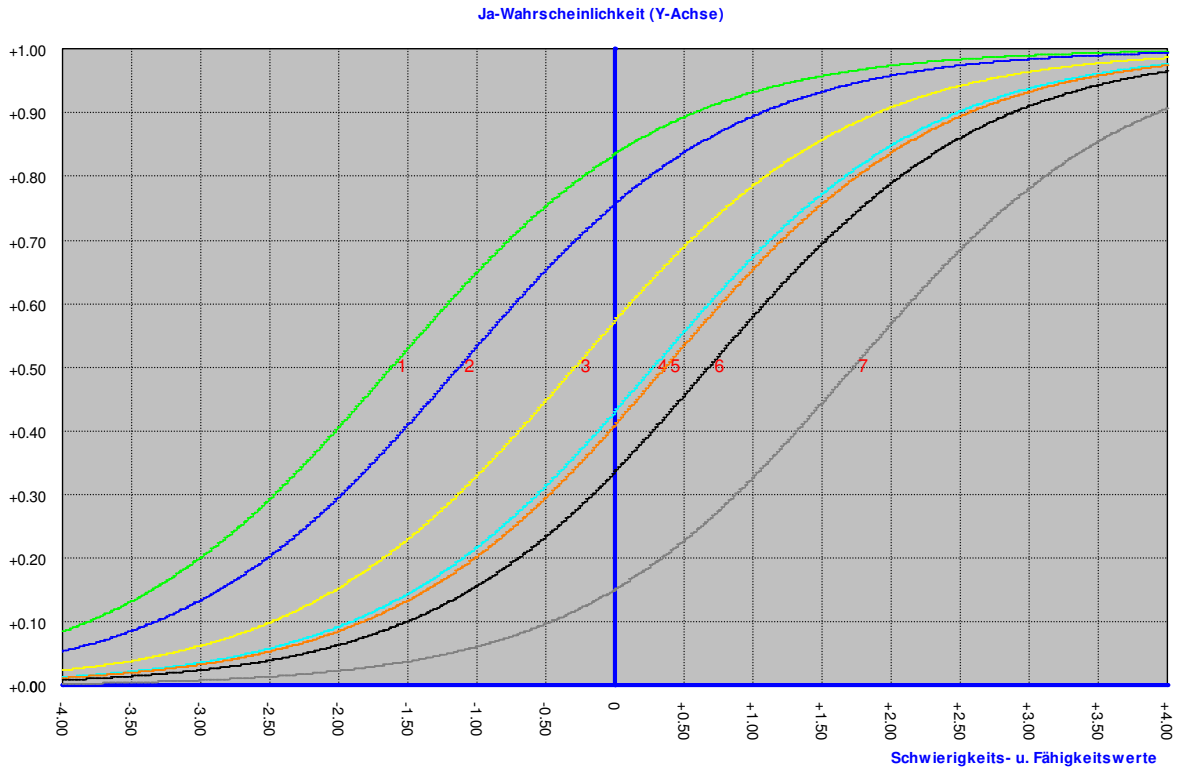


```
Ogive Y=Integral(1.259/2.51 * e**(-(1.259*(X-0.157))*(1.259*(X-0.157))/2))
rote Punkte=empirische Ja-Wahrscheinlichkeit bei einem Gesamtpunktwert von x
Die Ges.punktwerte stehen bei den Punkten. Ges.punktwerte in Klammern geschätzt
```

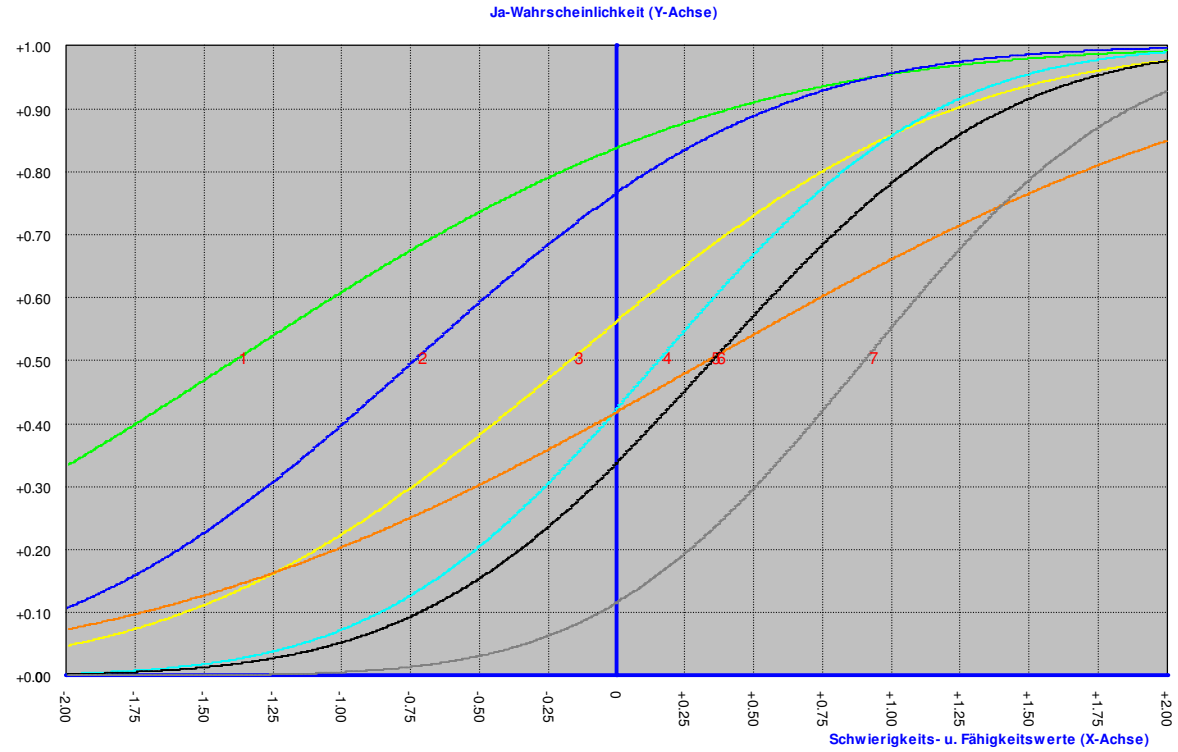



Alle Kurven in einer Grafik gezeichnet sind dann für Rasch und LSA

Logistische Funktionen




Ogiven



Aus der LSA erhalten wir folgende Trennschärfen

| | | Trennschaerfe (=Steilheit) | |
|---|---|---|---|
| V1 | M033Q01.MATH.P2000 | 0.709 | |
| V3 | M155Q01.MATH.P2000 | 0.984 | |
| V4 | M155Q04T.MATH.P2000 | 0.914 | |
| V5 | M192Q01T.MATH.P2000 | 1.259 | |
| V6 | M406Q01.MATH.P2003 | 1.196 | |
| V7 | M406Q02.MATH.P2003 | 1.327 | |
| V9 | M411Q02.MATH.P2003 | 0.620 | *** |

Bei der Grafik aus der LSA ist deutlich zu erkennen, dass die orange Kurve, das ist V9, quer durch die Kurven von V4,5,6,7 schneidet. Sie hat mit 0.620 die geringste Steilheit. Die anderen 6 Kurven schneiden sich nicht, bzw. nur geringfügig im oberen Verlaufsbereich

Beim Rasch-Modell wird durch die Itemfit-Tests V9 als nicht-Raschkonform erkannt

Bei outfit erweist sich V9 als das schlechteste Item

| Variable | | outfit | t *) | |
|---|---|---|---|---|
| V1 | M033Q01.MATH | 1.12709 | 0.93263 | |
| V3 | M155Q01.MATH | 0.97953 | -0.15343 | |
| V4 | M155Q04T.MATH | 1.07264 | 0.92489 | |
| V5 | M192Q01T.MATH | 0.91556 | -1.11802 | |
| V6 | M406Q01.MATH | 0.92997 | -0.85382 | |
| V7 | M406Q02.MATH | 0.74843 | -2.13753 | ? |
| V9 | M411Q02.MATH | 1.22678 | 2.83389 | ?? |

*) ein t-Wert zwischen -1.96 und +1.96 bedeutet: Die Variable ist "Rasch-konform"
ein nicht-Rasch-konformer Item wird durch ein Fragezeichen ? markiert

Beim Q-Index (nach Rost/von Davier) hat V9 den grössten Q-Wert, also die geringste Steilheit. Die Differenz zum Rasch-perfekten Q-Idealwert ist für V9 am grössten.

| Variable | | Q-Index | Rasch-perfekter Idealwert |
|---|---|---|---|
| V1 | M033Q01.MATH | 0.15912 | 0.10652 |
| V3 | M155Q01.MATH | 0.10234 | 0.10780 |
| V4 | M155Q04T.MAT | 0.13588 | 0.10504 |
| V5 | M192Q01T.MAT | 0.09069 | 0.10601 |
| V6 | M406Q01.MATH | 0.09163 | 0.10496 |
| V7 | M406Q02.MATH | 0.05633 | 0.10462 |
| V9 | M411Q02.MATH | 0.19011 ?? | 0.10757 |

Der Q-Wert von V9 weicht stark vom Idealwert ab.

Die Trennschärfe aus der LSA und der Q-Index aus dem Rasch-Modell stimmen sehr gut (gegenläufig) überein

| | | LSA Trennschaerfe (=Steilheit) | | Rasch Q-Index | |
|---|---|---|---|---|---|
| V1 | M033Q01.MATH.P2000 | 0.709 | | 0.15912 | |
| V3 | M155Q01.MATH.P2000 | 0.984 | | 0.10234 | |
| V4 | M155Q04T.MATH.P2000 | 0.914 | | 0.13588 | |
| V5 | M192Q01T.MATH.P2000 | 1.259 | | 0.09069 | |
| V6 | M406Q01.MATH.P2003 | 1.196 | | 0.09163 | |
| V7 | M406Q02.MATH.P2003 | 1.327 | | 0.05633 | |
| V9 | M411Q02.MATH.P2003 | 0.620 | *** | 0.19011 | ?? |

Beim Andersen-Test ist V3 das schlechteste Item und V9 das 2. schlechteste

| Item | Distanz zu Diagonale | mittlere Distanz zur Diagonalen aus 200 Simulationen | bei Simulationen war Distanz groesser Faelle | p | |
|---|---|---|---|---|---|
| V1 | 0.2677 | 0.1276 | 21 | 0.1050 | |
| V3 | 0.8928 > | 0.1284 | 0 | 0.0000 | ? |
| V4 | 0.0898 | 0.1353 | 120 | 0.6000 | |
| V5 | 0.1307 | 0.1270 | 87 | 0.4350 | |
| V6 | 0.0916 | 0.1270 | 115 | 0.5750 | |
| V7 | 0.1386 | 0.1198 | 70 | 0.3500 | |
| V9 | 0.7130 > | 0.1148 | 0 | 0.0000 | ? |

Wir haben oben bereits darauf hingewiesen, dass eine Itemkurve, die durch andere schneidet, bei der "Mokken-Analyse" über den H-Koeffizienten identifiziert werden kann. Mit Prog16m1 (erreichbar über den Knopf "Verfahren", dann "Skalierungsverfahren") haben wir eine Mokken-Analyse der 7 Pisa-Mathe-Items gerechnet und folgende H-Werte erhalten

Mokken-Analyse

| Variable | H | |
|---|---|---|
| V1 M033Q01 | 0.256 | |
| V3 M155Q01 | 0.366 | |
| V4 M155Q04 | 0.294 | |
| V9 M411Q02 | 0.207 | ?? |
| V5 M192Q01 | 0.367 | |
| V6 M406Q01 | 0.382 | |
| V7 M406Q02 | 0.572 | |

V9 besitzt den schlechtesten H-Wert. Mokken empfiehlt, Items nur bis zu einem H-Wert von ca. 0,3 zu akzeptieren. Dann müsste zumindest auch V1 ausgeschlossen werden. Siehe zur Mokken-Analyse das Almo-Dokument 19

Joachim Gerich: Guttman- und Mokken-Skalierung

Abschliessend haben wir noch die Fähigkeitswerte für alle Personen aus dem Rasch-Modell und aus dem LSA-Ogiven-Modell ermittelt und diese beiden korreliert. Es entstand ein Korrelationskoeffizient von r = 0.9957. Das bedeutet, dass die Fähigkeitswerte des einen Verfahrens als lineare Transformation des anderen Verfahrens betrachtet werden können - also identisch sind.
LSA-Ogiven- und Rasch-Modell sind austauschbar.

Unsere **Empfehlung** ist: Sind die Items dichotom, dann ist es für das Rasch-Modell empfehlenswert als (sozusagen) zusätzlichen Itemfit-Test ein LSA-Ogiven-Modell und eine Mokken-Analyse zu rechnen. Dadurch werden Items mit abweichender Steilheit sehr gut erkannt. Allerdings ist dieser "zusätzliche Itemfit-Test" nur auf dichotome Items anwendbar.

## Literatur zum ordinalen Rasch-Modell:

Ausführliche Literaturliste bei Rost, 2004


M. Bühner: Einführung in die Test- und Fragebogenkonstruktion, München Pearson 2011

Hermann Denz: Analyse latenter Strukturen, UTB, 1198, 1982

Hermann Denz: Latent structure analysis, Almo-Dokument 20, im Internet herunterladbar

G.H. Fischer & I.W. Molenaar: Rasch-Models – foundations, recent developments and applications, 1995, Springer, New York

Joachim Gerich: Guttman- und Mokken-Skalierung, in Wolf/Best: Handbuch der sozialwissenschaftlichen Datenanalyse, Springer Verlag, 2010

Joachim Gerich: Guttman- und Mokken-Skalierung, Almo-Dokument 19, im Internet herunterladbar

Moritz Heene1, Clemens Draxler, Matthias Ziegler & Markus Bühner:
Performance of the bootstrap Rasch model test under violations of non-intersecting item response functions, Psychological Test and Assessment Modeling, Volume 53, 2011 (3), 283-294, (Behandelt Cressie-Read-Test. Kann im Internet herunter geladen werden)

I.Koller, R.Alexandrowicz, R.Hatzinger: Das Rasch Modell in der Praxis, eine Einführung mit eRm, UTB, 2012

Clemens Lechner und Moritz Niehaus: Skript zur Vorlesung *Theorien psychometrischer Tests II,* WS 2009/2010, Dozent: Norman Rose, beschreibt das Programm Winmira, im Internet

J. A. López Pina, M. D. Hidalgo Montesinos: Fitting Rasch Model using Appropriateness Measure Statistics, The Spanish Journal of Psychology 2005, Vol. 8, No. 1, 100-110 1138-7416, im Internet herunterladbar

G.N. Masters: A Rasch-Model for partial credit scoring, Psychometrica, 47, S. 149-174

Muraki, E.: A generalized partial credit model: application of an EM algorithm. Applied Psychological Measurement, 16, 1992, S.159–176.

Jürgen Rost: Lehrbuch Testtheorie - Testkonstruktion, Verlag Hans Huber, 2004

Jürgen Rost, Matthias von Davier: A conditional item-fit index for Rasch-models

Applied psychological measurement, Vol. 18, June 1994, S.171-182

Friedrich Sixtl: Skalierungsverfahren, in Holm: Befragung 4, UTB, 434

Carolin Strobl: Das Rasch-Modell: Eine verständliche Einführung, Rainer Hamp Verlag, München, 2010

B. Wright, G.N. Masters: Rating Scale Analysis, 1982, S.100, Formel-Sammlung in www.rasch.org/rmt/rmt34e.htm

B. Wright & M. Linacre: Reasonable mean-square fit values, www.rasch.org/rmt/rmt83b.htm