

Die Pairwise-Methode zur Parameterschätzung im ordinalen Rasch-Modell

Jörg-Henrik Heine¹ & Christian Tarnai²

¹Ludwig-Maximilians Universität München

¹Technische Universität München

²Universität der Bundeswehr Neubiberg (bei München)

20. September 2013



Technische Universität München

Drei Standard Schätzmethoden zur Bestimmung von Item (und Personenparametern) im Raschmodell

1 CML:

- Getrennte Schätzung von Item- und Personenparametern
- Unabhängig von der Score Verteilung
- Problematisch bei hohen missing Anteilen (z.B. Booklet-Design)

2 MML:

- Getrennte Schätzung von Item- und Personenparametern
- Unproblematisch bei Booklet-Design Datensätzen (missing by design)
- Problematisch bei Verletzung der Normalverteilungsannahme (z.B. kleine Datensätze, selektive Stichproben)

3 JML:

- Gleichzeitige Schätzung von Item- und Personenparametern
- Personenparameter Schätzer stets ungenauer als Itemparameter Schätzer

Allgemein ist bei mehrstufigen Antwortformaten und kleinen Datensätzen mit geringen Kategoriehäufigkeiten zu rechnen, was in der Regel zu Schätzproblemen führt.

- *„These conditional methods of parameter estimation for items with more than two ordered categories involve sufficient statistics for the threshold parameters that involve directly the frequencies of responses in the adjacent categories. This is also a feature of other methods of estimation, for example, the marginal maximum likelihood (Bock and Aitkin, 1981) and the joint maximum likelihood (Wright and Douglas, 1977) methods. As a result, and in the case of some dislocation of persons and items which can result in **some categories of some items having low frequencies, the estimates of corresponding thresholds will be unstable.**“* (Andrich & Luo, 2003)

Explizite Berechnung der Itemparameter nach der Methode des „paarweisen Itemvergleichs“ und anschließende ML-Schätzung der Personenparameter

① zentrale Literatur zu Pairwise:

- Entwicklung der Methode zur Itemkalibrierung in Itembanken (Choppin, 1968, 1983)
- Generalisierung auf polytome Items (Garner & Engelhard Jr, 2002; Wright & Masters, 1982)

② Prinzip:

- Zählen der „Lösungs-“Häufigkeiten von Item j unter Voraussetzung das Item i nicht gelöst wurde
- Bildung der negativ reziproken Häufigkeits-Matrix
- Logarithmieren der Zeilenrandsummen
- Detailliertere Darstellung (für dichotome Daten) vgl. (Heine, Tarnai & Hartmann, 2011)

① Vorteile:

- Getrennte Schätzung von Item- und Personenparametern
- „*the method is simpler to implement than the CML*“ (Andrich & Luo, 2003)
- Konsistenz nachgewiesen für dichotome Items (Zwinderman, 1995)
- Und schließlich . . .
„*A special advantage of PAIR is that it can be used to analyze the incomplete data matrices which result when some items are not taken by some persons in the calibration sample.*“
(Wright & Masters, 1982, S. 69)

1 Versions Historie:

- Erste Version (0.1.3) auf CRAN verfügbar seit 11. Februar 2013
- Aktuelle Version ist die Version (0.1.4; seit 11. Juli 2013) (Heine, 2013)
- Kommende Version (0.1.5) als vorab Paket unter:
www.joerg-henrik-heine.de

2 Funktionsumfang:

- Itemparameter Berechnung für dichotome und polytome Items
- Graphischer Modelltest (nach verschiedenen Splitting-Kriterien)
- Standardfehler Berechnung implementiert über bootstrap bzw. jack-knive
- S3-methods für Ergebnisobjekte: `summary()` und `plot()`

3 geplanter Funktionsumfang:

- Integration der Personenparameter Schätzung
- Ggf. weitere Modelltests . . .
- Zusammenfassung der einzelnen Funktionen in einer zentralen wrapper Funktion (Anwenderfreundlichkeit)

① **Verlässlichkeit der Parameterschätzung:**

- Wie gut lassen sich die „*wahren*“ Itemparameter mit {pairwise} anhand von unvollständigen Daten bestimmen?
- Wie schneiden andere Methoden im Vergleich dazu ab?
- Wie groß sind die zu erwartenden Abweichungen bei der Parameterschätzung mit unvollständigen Daten im Vergleich zu vollständigen Daten?

② **Standardfehler:**

- Wie gut lassen sich die Standardfehler mit {pairwise} bestimmen ...
- ... bei unvollständigen Daten ?
- Wie schneiden andere Methoden im Vergleich dazu ab?

① Stichprobenbeschreibung:

- Männliche Studierende der Universität der Bundeswehr (1999 - 2002)
- Erhebung im Rahmen der Methoden Ausbildung im Studiengang Diplom Pädagogik.
- $n = 620$ (Vollständige Daten)

② Skalenbeschreibung:

- 8 Items der NEO-PI-R Facette „Befangenheit“ (N4)
- Das ursprünglich 5–Stufige Antwortformat wurde durch zusammenlegen der 3. und 4. Kategorie auf 4 Stufen reduziert

③ Erzeugung fehlender Werte:

- Erzeugen von (künstlich) fehlenden Werten in der empirischen Datenmatrix
- Ansteigender prozentualer Anteil an „missings“: in 5% Schritten von 5% bis 40%.
- Zufälliger Ausfallmechanismus

pol	nam	text
+	N017	Im Umgang mit anderen befürchte ich häufig, daß ich unangenehm auffallen könnte.
+	N047	Manchmal war mir etwas so peinlich, daß ich mich am liebsten versteckt hätte.
-	N077	Es bringt mich nicht besonders in Verlegenheit, wenn andere mich verspotten und lächerlich machen.
-	N107	Ich bin selten verlegen, wenn ich unter Leuten bin.
+	N137	Ich fühle mich anderen oft unterlegen.
-	N162	In Gegenwart meiner Chefs oder anderer Autoritäten fühle ich mich wohl.
+	N167	Wenn ich einer Person etwas falsches gesagt oder angetan habe, kann ich es kaum ertragen, ihr noch einmal zu begegnen.
+	N217	Wenn meine Bekannten dummen Unfug treiben, so ist mir das peinlich.

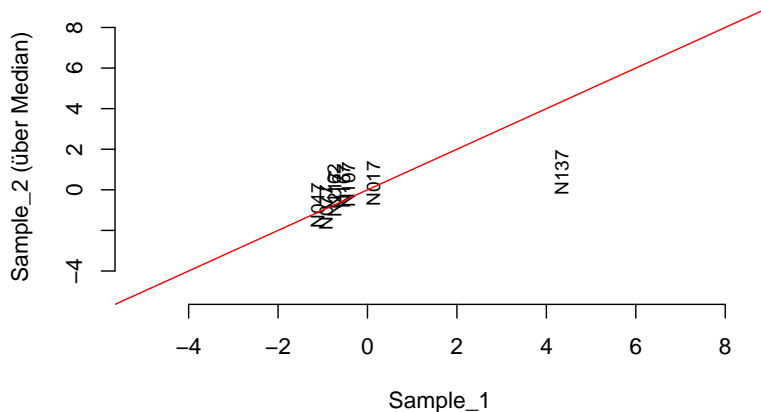
① Daten (Baseline):

- Aufbereitung der Daten Entfernung von „unit–non–respondern“ und Personen mit konstanten Antwortvektoren
–> vollständige Datenmatrix
- Prüfung der Modellgeltung (PCM) mit **WINMIRA**, **{pairwise}** und **ALMO**
- Bestimmung der Itemparameter als Baseline für Vergleich mit ...

② Auswertung der Datensätze mit fehlenden Werten:

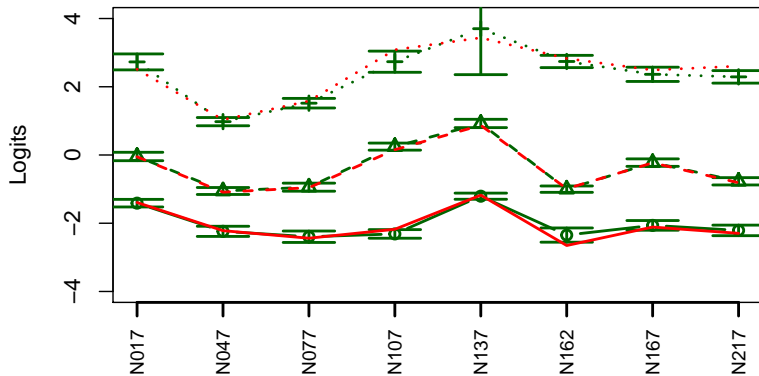
- Bestimmung der Itemparameter bei zunehmendem Anteil fehlender Werte ...
 - ... mit CML–Schätzung (WINMIRA; von Davier, 2001)
 - ... mit PAIR–Berechnung (**{pairwise}**; Heine, 2013)
 - ... mit Imputation mit SPSS und CML–Schätzung (WINMIRA; von Davier, 2001)
 - ... mit MML–Schätzung (ACER–ConQuest; Wu, Adams, Wilson & Halda, 2012)
- Vergleich des Verlaufs der Abweichung der jeweiligen Punktschätzer von den Punktschätzern auf Basis des vollständigen Datensatzes
- Vergleich des Veränderung der Standardfehler

pairwise: Modelltest Mediansplit; 0 % missing



Abweichung CML vs {pairwise}

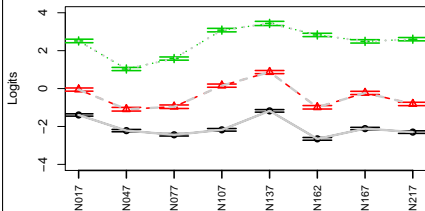
CML vs PAIR: 0 % missing; (n = 620)



rote Linien: CML Parameter Schätzung mit WINMIRA
grüne Linien: PAIR Parameter Berechnung mit {pairwise}

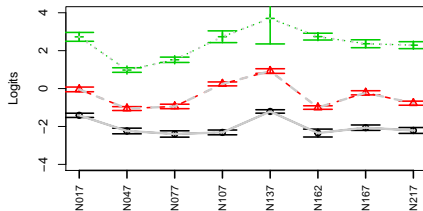
Itemparameter-Schätzung bei fehlenden Werten

CML: 0 % missing; (n = 620 ; listwise)



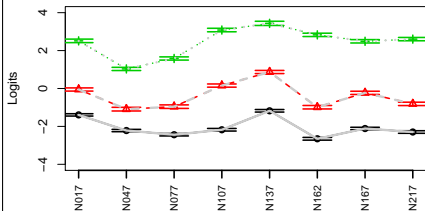
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 0 % missing; (n = 620)



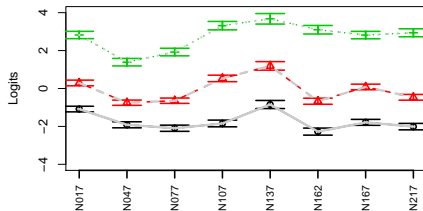
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 0 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

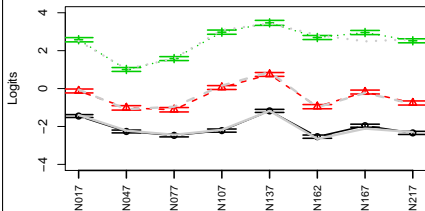
MML: 0 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

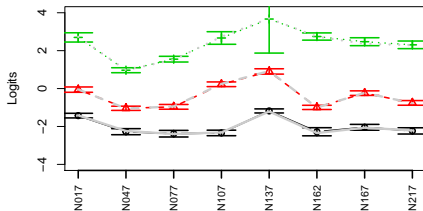
Itemparameter-Schätzung bei fehlenden Werten

CML: 5 % missing; (n = 418 ; listwise)



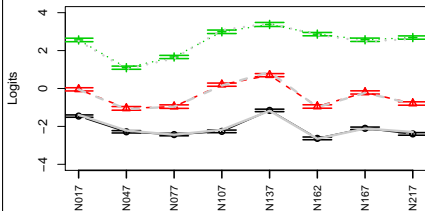
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 5 % missing; (n = 620)



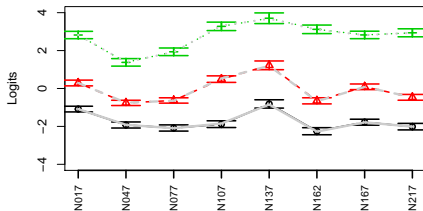
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 5 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

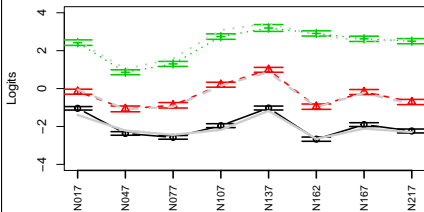
MML: 5 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

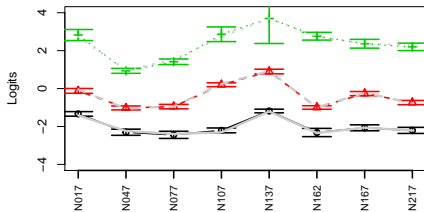
Itemparameter-Schätzung bei fehlenden Werten

CML: 10 % missing; (n = 255 ; listwise)



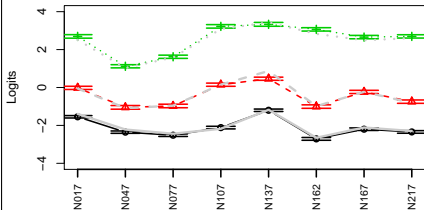
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 10 % missing; (n = 620)



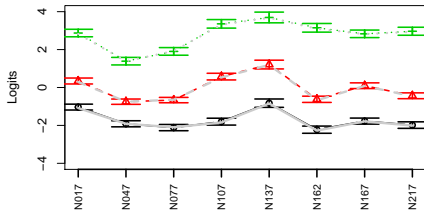
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 10 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

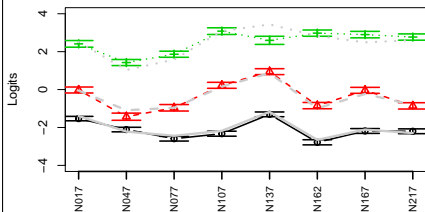
MML: 10 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

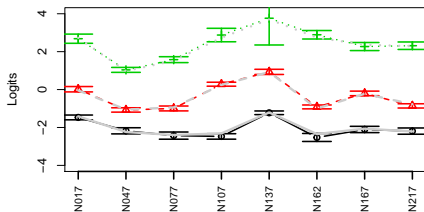
Itemparameter-Schätzung bei fehlenden Werten

CML: 15 % missing; (n = 176 ; listwise)



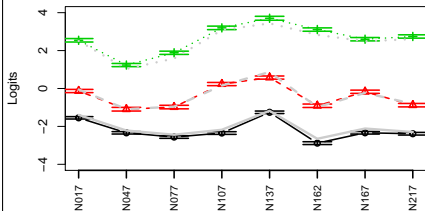
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 15 % missing; (n = 620)



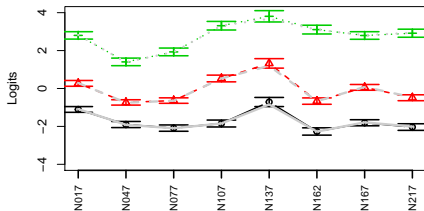
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 15 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

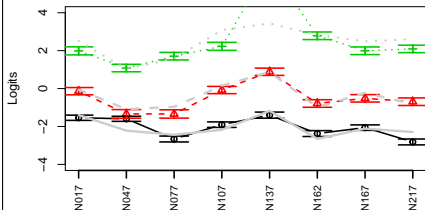
MML: 15 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

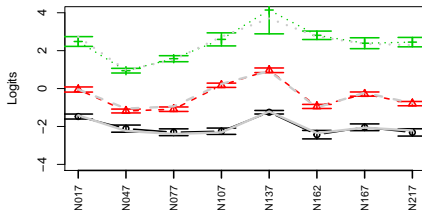
Itemparameter-Schätzung bei fehlenden Werten

CML: 20 % missing; (n = 117 ; listwise)



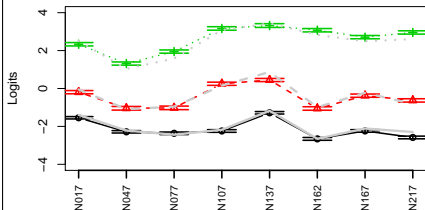
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 20 % missing; (n = 620)



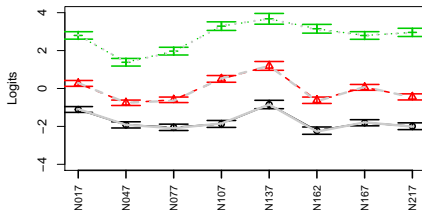
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 20 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

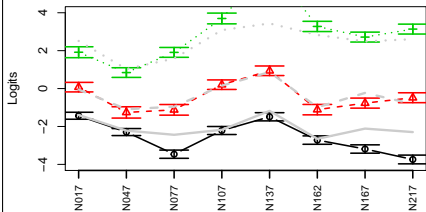
MML: 20 % missing; (n = 620)



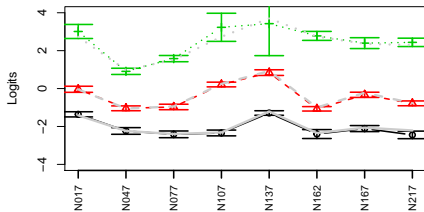
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

Itemparameter-Schätzung bei fehlenden Werten

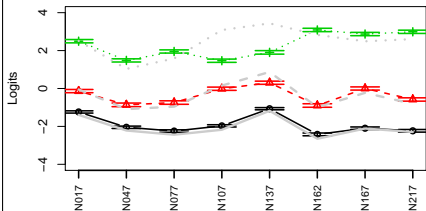
CML: 25 % missing; (n = 69 ; listwise)



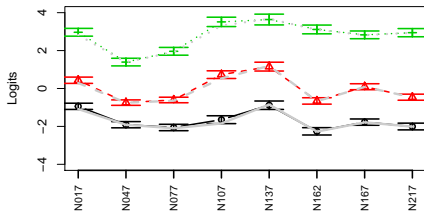
PAIR: 25 % missing; (n = 620)



CML imputiert: 25 % missing; (n = 620)

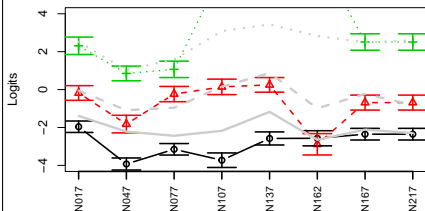


MML: 25 % missing; (n = 620)



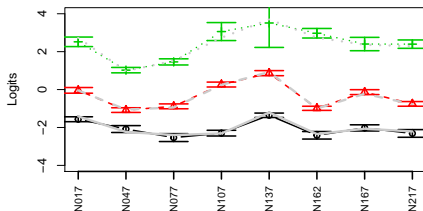
Itemparameter-Schätzung bei fehlenden Werten

CML: 30 % missing; (n = 30 ; listwise)



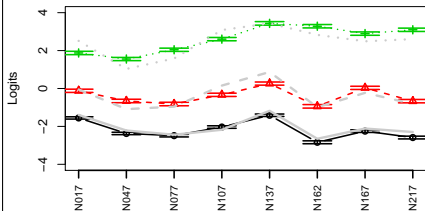
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 30 % missing; (n = 620)



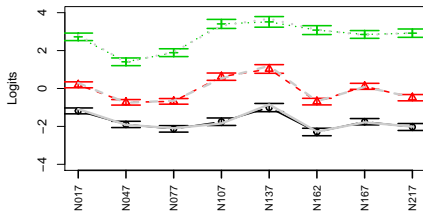
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 30 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

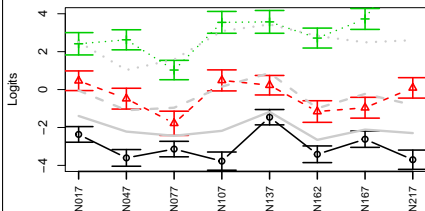
MML: 30 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

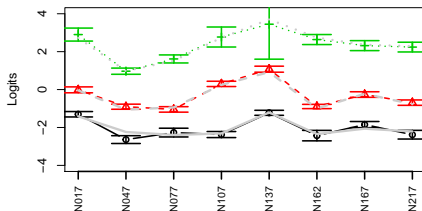
Itemparameter-Schätzung bei fehlenden Werten

CML: 35 % missing; (n = 19 ; listwise)



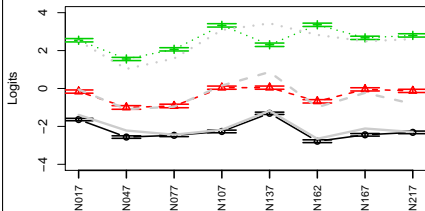
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 35 % missing; (n = 620)



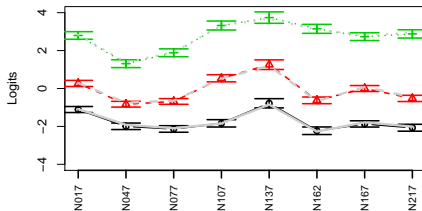
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 35 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

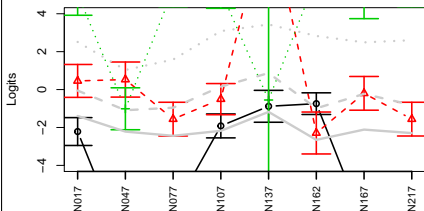
MML: 35 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

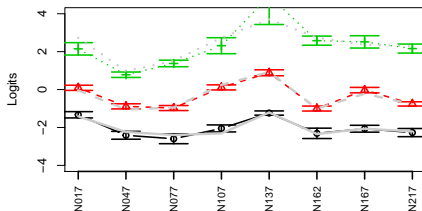
Itemparameter-Schätzung bei fehlenden Werten

CML: 40 % missing; (n = 6 ; listwise)



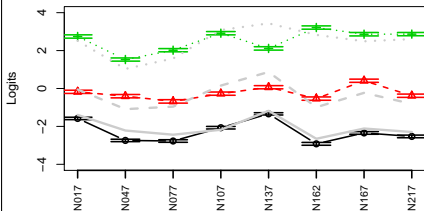
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

PAIR: 40 % missing; (n = 620)



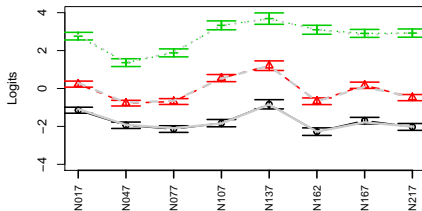
graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

CML imputiert: 40 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

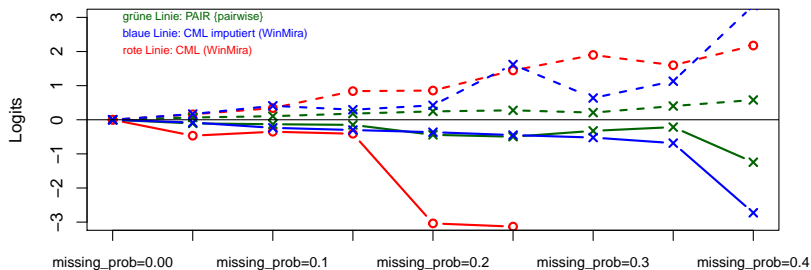
MML: 40 % missing; (n = 620)



graue Linien: Parameter Schätzung auf Basis original Daten; (n=620)

Range der Abweichung Itemparameter-Schätzung

Min / Max Abweichung Itemthresholds über jeweils 8 Items



- 1 **Standardfehler:** Mit zunehmendem Anteil fehlender Werte ...
 - ... kontinuierliche Vergrößerung der Standardfehler mit CLM-Schätzung
 - ... bleibt die Größe der Standardfehler mit {pairwise} und MML-Schätzung stabil
- 2 **Punktschätzer:** Mit zunehmendem Anteil fehlender Werte ...
 - ... werden die Abweichung von den ursprünglich geschätzten Werten bei CLM-Schätzung größer
 - ... verbessert die Imputation die CLM-Schätzung nur bedingt
 - ... steigen die Abweichungen von den ursprünglich berechneten Werten mit {pairwise} und MML-Schätzung nicht so stark an
- 3 **Kritik:** {pairwise} Standardfehler bei wenig fehlenden Daten größer
- 4 **Fazit für die Praxis:**
 - Itemparameterschätzung mit {pairwise} auch bei „löchrigen“ Daten möglich → (relative) Stabilität der Punktschätzer
 - {pairwise} als gute Alternative zu MML (kommerzielles Programm ConQuest).
 - Frage: Relevanz von größeren Standardfehlern mit {pairwise} – insbesondere bei Skalenentwicklung bzw. Pretest-Daten

- Andrich, D. & Luo, G. (2003). Conditional pairwise estimation in the rasch model for ordered response categories using principal components. *Journal of Applied Measurement*, 4 (3), 205–221.
- Choppin, B. (1968). Item bank using sample-free calibration. *Nature*, 219 (5156), 870–872.
- Choppin, B. (1983). *a fully conditional estimation procedure for rasch model parameters* (CSE Report Nr. 196).
- Garner, M. & Engelhard Jr, G. (2002). An eigenvector method for estimating item parameters of the dichotomous and polytomous rasch models. *Journal of Applied Measurement*, 1 (3), 107–128.
- Heine, J.-H. (2013). *pairwise: Rasch model parameters by pairwise algorithm (v 0.1.4)*. Zugriff auf <http://cran.r-project.org/web/packages/pairwise/index.html> (R package version 0.1.2)
- Heine, J.-H., Tarnai, C. & Hartmann, F. G. (2011). Eine methode zur parameterbestimmung im rasch-modell bei fehlenden werten. Bamberg. (21.09.2011 - 23.09. 2011)
- von Davier, M. (2001). *WINMIRA 2001*. Groningen, The Netherlands: ASC-Assessment Systems Corporation USA and Science Plus Group.
- Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Wu, M., Adams, R. J., Wilson, M. & Halda. (2012). *ACER ConQuest: generalised item response modeling software*. Melbourne: ACER.
- Zwinderman, A. H. (1995). Pairwise parameter estimation in rasch models. *Applied Psychological Measurement*, 19 (4), 369–375.