

Tests der Annahmen des Rasch-Modells

Seminar Psychometrische Modelle: Theorie und Anwendungen

Ender Celik

Institut für Statistik
Ludwig-Maximilians-Universität München

Juni 2014

Übersicht

1 Modellannahmen

- Eindimensionalität
- Lokale stochastische Unabhängigkeit
- Spezifische Objektivität

2 Parametrische Testverfahren

- Der bedingte Likelihood-Quotienten-Test
- Der Wald-Test
- Graphische Modellkontrolle
- Der Martin-Löf-Test

3 Nichtparametrische Testverfahren

- Motivation
- Quasi exakte Tests beim Rasch-Modell
- Test nach Ponocny

Übersicht

- 1 Modellannahmen
- 2 Parametrische Testverfahren
- 3 Nichtparametrische Testverfahren

Modellannahmen

Das Rasch-Modell (RM) (Rasch,1960)

Die Modellgleichung für das Rasch-Modell lautet :

$$P(X_{vi} = x_{vi} | \theta_v, \beta_i) = \frac{\exp(x_{vi}(\theta_v - \beta_i))}{1 + \exp(\theta_v - \beta_i)} \quad (1)$$

- θ_v : Personenfähigkeitsparameter für $v = 1, \dots, n$
- β_i : Itemschwierigkeitsparameter für $i = 1, \dots, k$
- x_{vi} : Antwort der v -ten Person auf das i -te Item

Modellannahmen

In der folgenden Datenmatrix, sind die Antworten von vier Personen auf drei Items dargestellt:

Personen	Items			$\sum_i x_{vi} = r_v$
	I_1	I_2	I_3	
P_1	1	0	1	2
P_2	1	0	0	1
P_3	1	1	0	2
P_4	1	1	1	3
$\sum_v x_{vi} = s_i$	4	2	2	-

Tabelle 1: Datenmatrix für vier Personen und drei Items

- Codierung: Item gelöst = 1 , Item nicht gelöst = 0
- r_v : Personenscores für jede einzelne Person $v = 1, \dots, 4$
- s_j : Itemscores für jedes Item $i = 1, \dots, 3$

Modellannahmen

Eindimensionalität

Eine zentrale Annahme des Rasch-Modells ist, die der Eindimensionalität, was bedeutet, dass die Personen- und Item-Parameter auf einer gemeinsamen latenten Dimension liegen.

$$P(X_{vi} = x_{vi} | \theta_v, \beta_i, \lambda) = P(X_{vi} = x_{vi} | \theta_v, \beta_i)$$

- Inhaltlich bedeutet dies, dass z.B. in einem Test zur mathematischen Kompetenz, nur die mathematische Kompetenz gemessen wird. Es spielt somit keine andere Fähigkeit, wie z.B. sprachliche Kompetenz, eine bedeutsame Rolle.

Lokale stochastische Unabhängigkeit

Die lokale stochastische Unabhängigkeit zwischen den Items besagt, dass bei Konstanthaltung der Personenfähigkeit θ_v die Itemantworten voneinander unabhängig sind.

$$\forall i, j : X_{vi} \perp X_{vj} | \theta_v$$

- Die lokale stochastische Unabhängigkeit ist zum Beispiel verletzt, wenn Aufgaben aufeinander aufbauen. Durch das Lösen des ersten Items, erhöht sich die Lösungswahrscheinlichkeit des zweiten Items.

Modellannahmen

Spezifische Objektivität

Die spezifische Objektivität bedeutet, dass Aussagen über die Personenfähigkeit von zwei Personen nicht davon abhängig sind, anhand welcher Items sie verglichen werden.

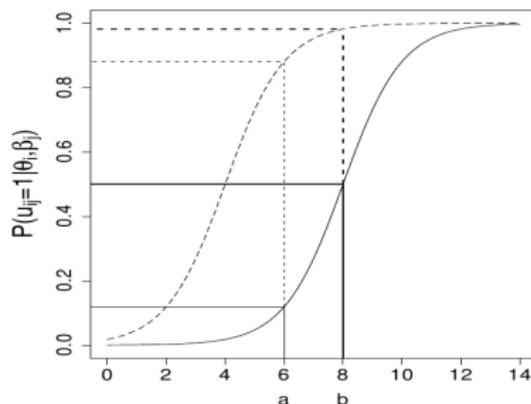


Abbildung 1: ICC: Vergleich der Lösungswahrscheinlichkeiten von zwei Personen, a und b, anhand von zwei Items (Quelle: Strobl 2012 , S.20)

Modellannahmen

Stichprobenunabhängigkeit

Die spezifische Objektivität wird auch als „**Stichprobenunabhängigkeit**“ bezeichnet. Wenn das Rasch-Modell für den Datensatz gilt, dann ist die Parameterschätzung der Personenfähigkeit und der Itemschwierigkeit unabhängig von der Itemstichprobe und der Personenstichprobe zu sehen. Die Genauigkeit der Schätzung hängt allerdings von der Stichprobengröße ab. Das bedeutet aber nicht, dass bei der Überprüfung auf Modellgültigkeit, keine Zufallsstichprobe benötigt wird.

- Stichprobenunabhängigkeit bedeutet nicht, dass z.B. ein Fragebogen zur Messung von Konfliktkompetenz für eine bestimmte Berufsgruppe direkt auf andere Berufsgruppen, übertragbar wäre. Personen aus anderen Berufsgruppen können die Fragen auch anders interpretieren (sog. **Differential Item Functioning (DIF)**). Deshalb muss für jede neue Gruppe von Personen, auf die der Test angewendet werden soll, erneut auf Modellgültigkeit überprüft werden.

Übersicht

- 1 Modellannahmen
- 2 Parametrische Testverfahren**
- 3 Nichtparametrische Testverfahren

Der bedingte Likelihood-Quotienten-Test

- Der bekannteste statistische Modelltest für das Rasch-Modell, ist der bedingte Likelihood-Quotienten-Test nach Andersen (1973).
- Ein Test auf Verletzungen der Eindimensionalität und spezifischen Objektivität (Differential Item Functioning).
- Der bedingte LQ-Test nach Andersen ist ein globaler Test. Das bedeutet, dass alle Items simultan geprüft werden.
- Die Grundidee: Teilt man die Personen in G Gruppen auf und schätzt für jede Gruppe getrennt die Itemschwierigkeitsparameter, sollte sich für jede Gruppe in etwa dasselbe ergeben.

Der bedingte Likelihood-Quotienten-Test

- Die bedingte Likelihood für den gesamten Datensatz:

$$L_c(\beta, \mathbf{r}) = \prod_{v=1}^n L_{c_v}(\beta, r_v) = \prod_{v=1}^n \frac{\exp(-\sum_{i=1}^k x_{vi}\beta_i)}{\gamma_{r_v}(\beta)} = \frac{\exp(-\sum_{i=1}^k s_i\beta_i)}{\prod_{v=1}^n \gamma_{r_v}(\beta)}$$

- Andersen hat eine Teilung des Datensatzes in $G = k - 1$ Gruppen nach dem Personenscore vorgeschlagen (z.B.: 4 Items, drei Gruppen: $r_v = 1, r_v = 2, r_v = 3$).

$$L_c^{(g)}(\beta, r^{(g)}) = \frac{\exp(-\sum_{i=1}^k s_i^{(g)}\beta_i)}{\gamma_{r_v}(\beta)^{n_g}}$$

- $s_i^{(g)}$: Itemscore für das Item i in der Personenscoregruppe g
- n_g : Anzahl der Personen in der Personenscoregruppe g

Der bedingte Likelihood-Quotienten-Test

- Die gesamte Likelihood ist das Produkt der Likelihoods für jede Personenscoregruppe:

$$L_c(\beta, \mathbf{r}) = \prod_{g=1}^G L_c^{(g)}(\beta, r^{(g)})$$

- Wenn alle Personen dieselben Itemparameter β haben, ist die Schätzung der Itemparameter $\hat{\beta}$, gemeinsam oder getrennt, gleich.

$$LQ_c = \frac{L_c(\hat{\beta}, \mathbf{r})}{\prod_{g=1}^G L_c^{(g)}(\hat{\beta}^{(g)}, r^{(g)})} = 1$$

Der bedingte Likelihood-Quotienten-Test

Der bedingte Likelihood-Quotienten-Test

H_0 : Itemparameter gleich, $\beta = \beta^{(1)} = \dots = \beta^{(G)}$

H_1 : mindestens zwei Itemparameter ($i \neq j$) unterschiedlich,

$\exists i, j : \beta^{(i)} \neq \beta^{(j)}$

$$T_{LQ_c} = -2 \ln \left(\frac{L_c(\hat{\beta}, \mathbf{r})}{\prod_{g=1}^G L_c^{(g)}(\hat{\beta}^{(g)}, r^{(g)})} \right) \sim \chi_{(k-1)(G-1)}^2$$

- In der Praxis ist eine Teilung in alle Personenscoregruppen nicht immer möglich, weil nicht in jeder Personenscoregruppe genügend Beobachtungen vorhanden sind → z.B.: Teilungskriterium der Median Personenscoreverteilung
- Bei der Überprüfung der *DIF*, ist das Teilungskriterium z.B. das Geschlecht oder die Berufsgruppen.

Der Wald-Test

- Ein Test auf Verletzungen der Eindimensionalität und spezifischen Objektivität (Differential Item Functioning).
- Der Wald-Test, ist ein Test auf Itemebene. Das bedeutet, dass einzelne Items geprüft werden.
- Im Gegensatz zum bedingten Likelihood-Quotienten-Test, werden die Schätzungen der Itemparameter direkt miteinander verglichen.
- Die Grundidee entspricht der beim bedingten Likelihood-Quotienten-Test.

Der Wald-Test

Der itemspezifische Wald-Test

H_0 : Itemparameter gleich, $\beta_i = \beta_i^{(1)} = \beta_i^{(2)}$

H_1 : Itemparameter nicht gleich, $\beta_i \neq \beta_i^{(1)} \neq \beta_i^{(2)}$

$$T_w = \frac{\hat{\beta}_i^{(1)} - \hat{\beta}_i^{(2)}}{\sqrt{\text{var}(\hat{\beta}_i^{(1)}) + \text{var}(\hat{\beta}_i^{(2)})}} \sim N(0, 1)$$

Graphische Modellkontrolle

- Teilung des Datensatzes in zwei Gruppen und Schätzung der Itemparameter der beiden Gruppen
- Die geschätzten Itemparameter der beiden Gruppen werden in einem Streudiagramm gegeneinander aufgetragen.
- Wenn das Rasch-Modell gilt, müssen die geschätzten Itemparameter in beiden Gruppen übereinstimmen.
- Graphisch betrachtet, sollten die geschätzten Itemparameter der beiden Gruppen auf der Winkelhalbierenden liegen.
- Unterscheiden sich die Gruppen hingegen systematisch, dann werden die geschätzten Itemparameter der beiden Gruppen weiter von der Winkelhalbierenden entfernt sein.
- Graphische Modellkontrollen sind rein deskriptiv und im eigentlichen Sinne kein Modelltest.

Graphische Modellkontrolle

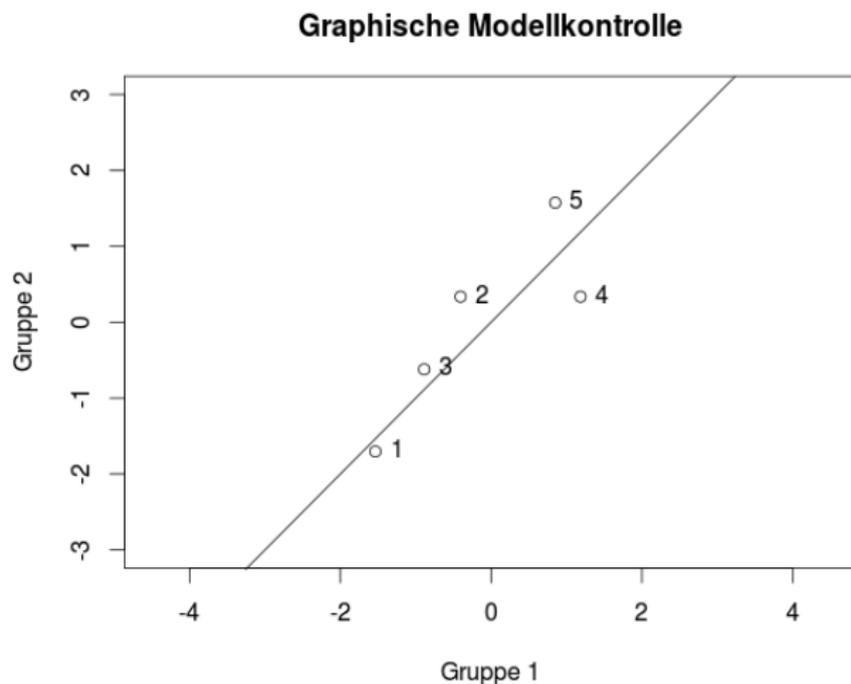


Abbildung 2: Streudiagramm für die geschätzten Itemparameter, getrennt in zwei Gruppen

Graphische Modellkontrolle

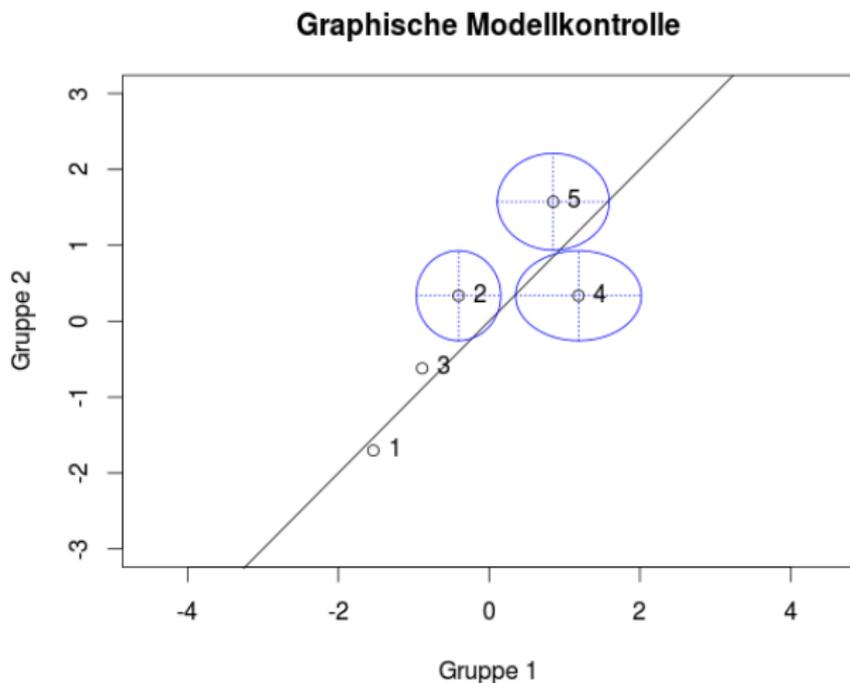


Abbildung 3: Streudiagramm für die geschätzten Itemparameter, getrennt in zwei Gruppen mit Konfidenzellipsen

Graphische Modellkontrolle

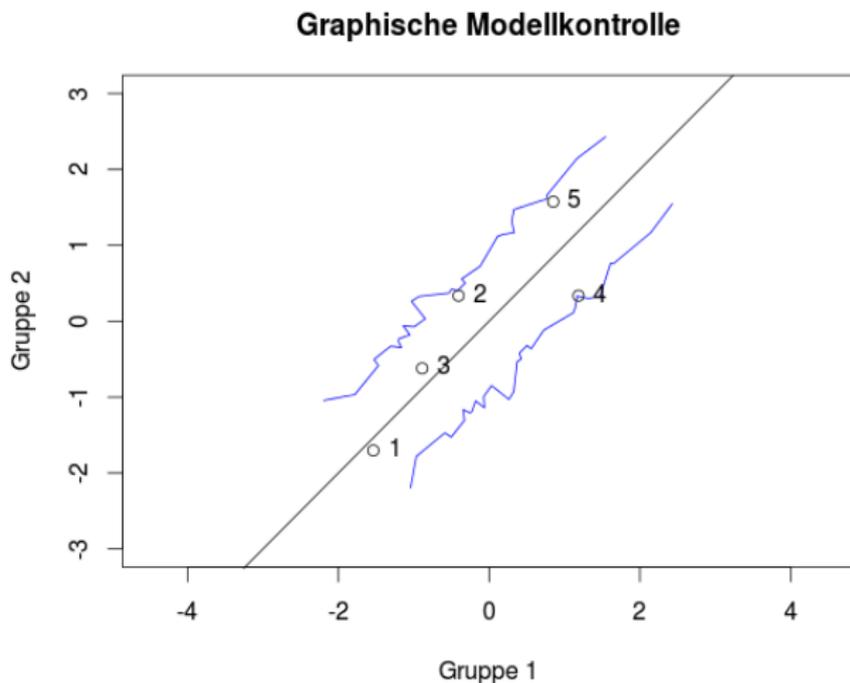


Abbildung 4: Streudiagramm für die geschätzten Itemparameter, getrennt in zwei Gruppen mit „punktweise Konfindenzintervalle“

Graphische Modellkontrolle

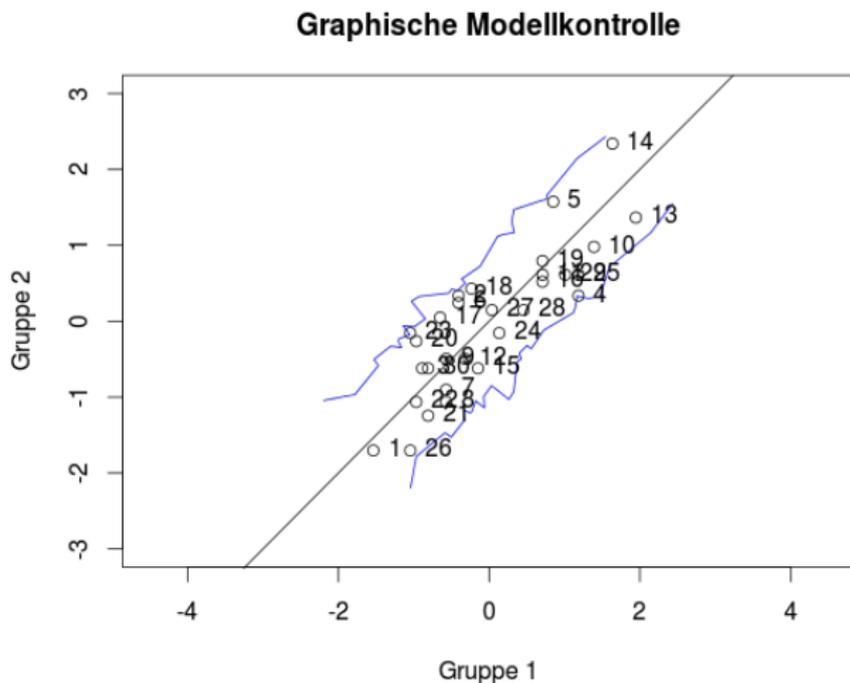


Abbildung 5: Streudiagramm für die geschätzten Itemparameter, getrennt in zwei Gruppen mit „punktweise Konfindenzintervalle“

Der Martin-Löf-Test

- Martin-Löf-Test (1973) auf Verletzungen der Eindimensionalität (Itemhomogenität)
- Im Gegensatz zu den vorherigen Tests, werden nicht die Personen, sondern die Items in Gruppen aufgeteilt.
- Die Grundidee: Teilt man die Items in 2 Gruppen auf, dann müssten die Personenscores für die beiden Itemgruppen in etwa gleich sein.

Der Martin-Löf-Test

- z.B.: Datensatz aus 4 Items und 54 Personen:

Personenscore r_v	0	1	2	3	4
Anzahl der Personen	3	9	25	10	7

Tabelle 2: Personenscoreverteilung der 54 Personen

- Teilung der Items in 2 Gruppen und Erstellung einer Kreuztabelle:

		lg 2			
		0	1	2	
lg 1	0	3	5	7	7 Personen zeigen das Antwortmuster {02}
	1	4	13	3	13 Personen zeigen das Antwortmuster {11}
	2	5	7	7	5 Personen zeigen das Antwortmuster {20}

Tabelle 3: Kreuztabelle der Itemgruppe 1 (lg 1) und Itemgruppe 2 (lg 2)

Der Martin-Löf-Test

- Beide Itemgruppen messen das Gleiche:

		lg 2		
		0	1	2
lg 1	0	18	0	0
	1	0	18	0
	2	0	0	18

Tabelle 4: Kreuztabelle der Itemgruppe 1 (lg 1) und Itemgruppe 2 (lg 2)

- Beide Itemgruppen messen etwas unterschiedliches:

		lg 2		
		0	1	2
lg 1	0	6	6	6
	1	6	6	6
	2	6	6	6

Tabelle 5: Kreuztabelle der Itemgruppe 1 (lg 1) und Itemgruppe 2 (lg 2)

Der Martin-Löf-Test

- Wenn beide Itemgruppen die gleichen Personenscores messen, dann werden sich die größten Häufigkeiten nahe der Hauptdiagonle befinden.
- Je weniger die Itemgruppen das gleiche messen, desto mehr Häufigkeiten werden von der Hauptdiagonale entfernt sein.
- Genügt ein Parameter für den Score $r_v = 2$, um die Häufigkeit von 25 zu beschreiben?
- Oder benötigt jedes Antwortmuster $\{02\}, \{11\}$ und $\{20\}$ einen eigenen Parameter?

Der Martin-Löf-Test

Der Martin-Löf-Test

H_0 : Beide Itemgruppen messen die dieselbe latente Eigenschaft

H_1 : Beide Itemgruppen messen unterschiedliche latente Eigenschaften

$$T_{ML} = -2 \ln \left(\frac{\prod_{r=1}^k \left(\frac{n_r}{n}\right)^{n_r}}{\prod_{r_1=1}^{k_1} \prod_{r_2=1}^{k_2} \left(\frac{n_{\{r_1 r_2\}}}{n}\right)^{n_{\{r_1 r_2\}}} \cdot \frac{L_c}{L_c^{(1)} \cdot L_c^{(2)}}} \right) \sim \chi^2_{(k_1 k_2 - 1)}$$

- $r = 0, \dots, k$: Personenscores im Gesamtdatensatz
- $r_1 = 0, \dots, k_1$: Personenscores in der ersten Itemgruppe (r_2 analog)
- n_r : Häufigkeiten der Personenscores im Gesamtdatensatz
- $n_{\{r_1 r_2\}}$: Häufigkeiten des Personenscores in der ersten und zweiten Itemgruppe
- $L_c, L_c^{(1)}, L_c^{(2)}$: bedingte Likelihoods, wie im Test nach Andersen

Übersicht

- 1 Modellannahmen
- 2 Parametrische Testverfahren
- 3 Nichtparametrische Testverfahren**

Motivation

- Die Asymptotik des χ^2 -Tests auf Unabhängigkeit bei Kontingenztafeln, greift nicht gut bei kleinen Stichproben ($n_{ij} \leq 5$).
- Lösung: „Exakter Test“ oder „quasi exakter Test“

	B	\bar{B}	Σ
A	3 (k)	7	10 ($n_{1.}$)
\bar{A}	8	8	16 ($n_{2.}$)
Σ	11 ($n_{.1}$)	15 ($n_{.2}$)	26 (n)

Tabelle 6: 2x2 Vierfeldertafel

- Unter der Annahme der Unabhängigkeit, ist bei gegebenen Randsummen, die Zelhäufigkeit n_{11} hypergeometrisch verteilt:

$$P(n_{11} = k) = \frac{\binom{n_{1.}}{k} \binom{n_{2.}}{n_{.1}-k}}{\binom{n}{n_{.1}}} = \frac{\binom{10}{3} \binom{16}{7}}{\binom{26}{11}} = 0.199$$

Motivation

- Der p-Wert ist die Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese, das vorhandene Ereignis oder „noch extremere“ Ereignisse zu beobachten.

$$P(n_{11} \leq 3) = P(n_{11} = 0) + P(n_{11} = 1) + P(n_{11} = 2) + P(n_{11} = 3) = 0.277$$

- Die Nullhypothese, dass beide Merkmale unabhängig voneinander sind, kann nicht verworfen werden.
- Bei $r \times c$ Kontingenztafeln mit großen Randsummen, ist die komplette Nummerierung aller Tafeln nicht möglich.
- Sind die Zellen dünn besetzt, ist auch die Verwendung des asymptotischen Tests fraglich.
- Lösung: quasi exakter Test (Monte Carlo-Test)

Motivation

quasi exakter Test

- 1 Ziehe B Kontingenztafeln KT_b ($b = 1, \dots, B$) aus der Menge aller Kontingenztafeln mit gegebenen Randsummen
- 2 Berechne Teststatistik $T_b = T(KT_b)$
- 3 Simulierter p-Wert ist Prozentsatz aller simulierten Tafeln, deren Teststatistik kleiner gleich (größer, . . .) als die Teststatistik T_0 der beobachteten Kontingenztafel KT_0 ist:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B t_b \quad \text{wobei} \quad t_b = \begin{cases} 1, & T_b \leq T_0 \\ 0, & \text{sonst} \end{cases}$$

Motivation

quasi exakter Test

```
urne=c(rep("A",10),rep("nA",16))
B <- 10000
T_0=3
t_b=0
for (b in 1:B) {
  sp=sample(urne, size=11, replace=F)
  T_b=sum(sp=="A")
  if (T_b <= T_0){
    t_b=t_b+1
  }
}
p=t_b/B; p
[1] 0.2797
```

Motivation

Vorteile

- Im Gegensatz zu den asymptotischen Tests, muss man nicht die asymptotische Verteilung der Teststatistik unter der Nullhypothese kennen.
- Funktioniert auch bei kleinen Stichproben
- Eignet sich bei großen Populationen, bei deren die Berechnung des exakten Tests sehr komplex wird.

Exakte Tests beim Rasch-Modell

- Bei Gültigkeit des Rasch-Modells, sind die Randsummen suffiziente Statistiken für die Personenfähigkeit und die Itemschwierigkeit.
- Die einzelnen Einträge in den Zellen sind irrelevant, solange die Randsummen gleich bleiben.
- Σ_{rs} ist die Menge aller möglichen Datenmatrizen mit gleichen r Zeilen- und s Spaltensummen.
- Für einen exakten Test müssen alle möglichen Datenmatrizen aufgestellt werden.
- Bereits für eine 12×12 große Matrix mit gleichen Randsummen $r = s = 2$, gibt es 21 959 547 410 077 200 mögliche Matrizen.
- Lösung: Zufallsstichprobe aus $\Sigma_{rs} \rightarrow$ quasi exakte Tests

Simulation von Matrizen

Anforderungen an die Simulation

- Die Ziehungen sollten
 - aus dem ganzen Stichprobenraum Σ_{rs} stammen
 - unabhängig sein
 - gleich wahrscheinlich sein

Der MCMC-Ansatz von Verhelst (2008)

Konstruktion einer Markov-Kette

- Alle möglichen Matrizen im Stichprobenraum $A \in \Sigma_{rs}$, werden als Zustände in einer endlichen Markov-Kette betrachtet.
- Die beobachtete Matrix wird als A_0 bezeichnet.
- Durch eine definierte Transformationsregel R , gelangt man mit nur einem Schritt von einer Matrix A_0 zu einer anderen Matrix A_1 im Stichprobenraum $A \in \Sigma_{rs}$.
- Alle Matrizen die mit dieser Transformation erreicht werden können, werden als „Nachbarschaft“ der Matrix A_0 bezeichnet $\mathcal{A}_R(A_0)$.

Der MCMC-Ansatz von Verhelst (2008)

Binomialtransformation und die „Binomiale Nachbarschaft“

Die Binomialtransformation B_{ij} ($i < j$) transformiert die i -te Spalte und j -te Spalte, alle anderen bleiben unverändert.

i1	i2	i3	i4
0	0	1	1
1	0	0	0
1	1	1	0
0	1	0	0
0	1	0	0

→

i2	i3
0	1
0	0
1	1
1	0
1	0

→

i2	i3
1	0
0	0
1	1
1	0
0	1

→

i1	i2	i3	i4
0	1	0	1
1	0	0	0
1	1	1	0
0	1	0	0
0	0	1	0

Tabelle 7: Binomialtransformation B_{23}

Der MCMC-Ansatz von Verhelst (2008)

- m_{ij} : Anzahl der Zeilen mit Zeilensumme eins, innerhalb des Spaltenpaares (i, j) ($m_{23} = 3$)
- a_{ij} : von m_{ij} die Spaltensumme in der i -ten Spalte ($a_{23} = 2$)
- $b_{ij} = m_{ij} - a_{ij}$: Anzahl der Einträge mit Null ($b_{23} = 1$)
- Binomialtransformation B_{ij} : Die a_{ij} Einsen der i -ten Spalte werden auf die m_{ij} verfügbaren Zeilen zufällig neu verteilt.
- Die restlichen b_{ij} Einträge der i -ten Spalte werden mit Nullen ergänzt.
- Die j -te Spalte wird so ergänzt, dass sich wieder jeweils die Zeilensumme 1 ergibt.
- Insgesamt existieren $\binom{m_{ij}}{a_{ij}} - 1$ Binomialtransformationen $\mathcal{A}_B^{ij}(A)$ der Matrix $A \in \Sigma_{rs}$.
- Die Binomiale „Nachbarschaft“ der Matrix $A \in \Sigma_{rs}$:

$$\mathcal{A}_B(A) = \bigcup_{(i,j)} \mathcal{A}_B^{ij}(A)$$

Der MCMC-Ansatz von Verhelst (2008)

- **Definition 1:** Das Spaltenpaar (i, j) heißt regulär (Guttman Paar), wenn:

$$a_{ij} \cdot b_{ij} > 0$$

- **Definition 2:** Das k_2 -Maß ist definiert für $A \in \Sigma_{rs}$ durch:

$$k_2(A) = \#\{(i, j) : i < j \leq k, \text{ ist reguläres Spaltenpaar von } A\}$$

- Übergangsmatrix:

$$\mathbf{P} = (p_{st}) = \frac{w_{st}}{k_2(A_s)}$$

mit

$$w_{st} = \begin{cases} \left(\binom{m_{ij}(A_s)}{a_{ij}(A_s)} - 1 \right)^{-1} & A_t \in \mathcal{A}_B^{ij}(A_s) \\ 0 & A_t \notin \mathcal{A}_B^{ij}(A_s) \end{cases}$$

- Stationäre Verteilung:

$$\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi} \text{ mit } \pi_t := k_2(A_t), (t = 1, \dots, \#\Sigma_{rs})$$

Der MCMC-Ansatz von Verhelst (2008)

- Stationäre Verteilung der Markov-Kette ist nicht gleichverteilt
- Lösung: Der Metropolis-Hastings-Algorithmus

Metropolis-Hastings-Algorithmus

- 1 Wähle zufällig ein reguläres Spaltenpaar (i, j) von A_s , aus der Menge $k_2(A_s)$ der regulären Spaltenpaare von A_s
- 2 Führe eine zufällige Binomialtransformation innerhalb des gewählten Spaltenpaares durch \rightarrow ergibt A_t
 - 1 Wenn $A_t = A_s$, dann wiederhole Schritt 2
 - 2 Akzeptiere A_t mit Wahrscheinlichkeit

$$\alpha_{st} = \min\left(1, \frac{p_{ts}}{p_{st}}\right) = \min\left(1, \frac{k_2(A_s)}{k_2(A_t)}\right)$$

- 3 Falls A_t akzeptiert wurde, ist der neue Zustand der Markov-Kette A_t , anderenfalls wird der alte Zustand beibehalten A_s

Der MCMC-Ansatz von Verhelst (2008)

- Burn-in: Die ersten simulierten Matrizen werden aus Gründen der Stationarität verworfen.
- Step: Die Abhängigkeit zwischen den Matrizen kann verringert werden, indem nicht jede k -te simulierte Matrix akzeptiert wird.

T_{11} – Test nach Ponocny

- Ein globaler Test auf Verletzungen der lokalen stochastischen Unabhängigkeit zwischen den Items und der Homogenität von Items.

T_{11} - quasi exakter Test

- 1 Ziehe B Datenmatrizen A_b ($b = 1, \dots, B$) mit dem MCMC-Sampler
- 2 Berechne Korrelation r_{ij} für jedes Itempaar
- 3 Berechne mittlere Inter-Itemkorrelationen $\tilde{r}_{ij} = \frac{1}{B} \sum_{b=1}^B r_{ij}^{(b)}$
- 4 Berechne Teststatistik $T_b = T(A_b) = \sum_{ij} |r_{ij} - \tilde{r}_{ij}|$
- 5 Simulierter p-Wert ist Prozentsatz aller zufällig gezogenen Datenmatrizen, deren Teststatistik größer gleich als die Teststatistik T_0 der beobachteten Matrix A_0 ist:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B t_b \quad \text{wobei} \quad t_b = \begin{cases} 1, & T_b \geq T_0 \\ 0, & \text{sonst} \end{cases}$$

T_{11} – Test nach Ponocny

- Überprüft, ob im gesamten Datensatz zu niedrige oder zu hohe Inter-Itemkorrelationen vorhanden sind.
- Der Test kann nicht unterscheiden, welche der Annahmen verletzt wurden.

T_1 – Test nach Ponocny

- Test auf Verletzungen der lokalen stochastischen Unabhängigkeit auf der Itemebene
- Überprüft, ob die Häufigkeit der Antwortmuster $\{00\}$ und $\{11\}$ überhöht ist.

T_1 -quasi exakter Test (für zwei Items)

- 1 Ziehe B Datenmatrizen A_b ($b = 1, \dots, B$) mit dem MCMC-Sampler
- 2 Berechne Teststatistik

$$T_b = T(A_b) = \sum_{v=1}^n \delta_{ij} \quad \text{wobei} \quad \delta_{ij} = \begin{cases} 1, & x_{vi} = x_{vj} \\ 0, & \text{sonst} \end{cases}$$

- 3 Simulierter p-Wert ist Prozentsatz aller zufällig gezogenen Datenmatrizen, deren Teststatistik größer gleich als die Teststatistik T_0 der beobachteten Matrix A_0 ist:

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B t_b \quad \text{wobei} \quad t_b = \begin{cases} 1, & T_b \geq T_0 \\ 0, & \text{sonst} \end{cases}$$

Literatur

- Koller, Hatzinger (2013): Nonparametric tests for the Rasch model: explanation, development, and application of quasi-exact tests for small samples, Interstat
- Strobl (2012): Das Rasch-Modell, Rainer Hampp Verlag
- Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rasch model. Psychometrika
- Jordan (2010): Nonparametrische Tests im Rasch-Modell - Teststärken unter verschiedenen Modellverletzungen
- Verhelst, Hatzinger, Mair (2007): The Rasch Sampler, Journal of Statistical Software
- Rost (2004): Testtheorie-Testkonstruktion, HH
- Koller, Alexandrowicz, Hatzinger (2012): Das Rasch Modell in der Praxis: Eine Einführung in eRm, UTB
- Chen, Y., & Small, D. (2005). Exact tests for the Rasch model via sequential importance sampling, Psychometrika