

Der Versuch eines Überblicks über psychometrische Modelle

Seminar: Statistische Modellierung latenter Strukturen in
den Lebens-, Sozial- und Wirtschaftswissenschaften

Institut für Statistik, LMU München

Can Gürer

18.01.2014

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Wie kann man Eigenschaften, Persönlichkeitsmerkmale (z.B. Intelligenz, Extraversion, Depression, Kompetenzen,...) messen/ messbar machen?
- ▶ Problem: Eigenschaften sind nicht direkt beobachtbar.
- ▶ Lösung: Schluss auf Eigenschaftsausprägungen über Verhalten.
 - ▶ Antworten auf Items als Manifestation von Verhalten.
 - ▶ Von diesen dann Schluss auf latente Merkmale.
- ▶ Das Ausgangsmaterial sind also Antworten von Personen auf viele Items zu einem Konstrukt.
- ▶ Auf welcher Basis soll der Schluss von Antwortverhalten auf Merkmalsausprägungen erfolgen? → Modellfrage

- ▶ KTT
- ▶ IRT (Personenvariablen kontinuierlich)
- ▶ CDM (Personenvariablen diskret)
- ▶ kombinierte Modelle

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe und Ansätze

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Güler

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

Klassische Testtheorie (KTT)

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Güler

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Ausgangsbasis:

$$X = T + E$$

- ▶ X beobachteter Wert
 - ▶ T wahrer Wert
 - ▶ E Fehler
- ▶ X z.B. Summenwert mehrerer Itemantworten.
- ▶ Annahmen:
 - $Erw(E) = 0$
 - $r(E, T) = 0$
 - $r(E_X, T_Y) = 0$ (Y : andere beobachtete Variable)
 - $r(E_X, E_Y) = 0$
- ▶ zentraler Kennwert ist die Reliabilität: $r_{tt} = \frac{var(T_X)}{var(X)}$
- ▶ "nur" Messfehlertheorie

IRT: Item Response Theory

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Gürer

Übersicht

KTT

IRT

Datenformat
Rasch-Modell
2PL-Modell
3PL-Modell
Andere Erweiterungen

CDM

DINA-Modell
Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Rasch Modell (1PL Modell)
- ▶ 2PL Modell
- ▶ 3PL Modell
- ▶ Erweiterungen (mehrdimensional, mehrkategorial)
- ▶ (Modelle mit Probit oder anderen Responsefunktionen)

Datenformat für die aufgeführten Beispielmodelle

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Gürer

Übersicht

KT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

► Response

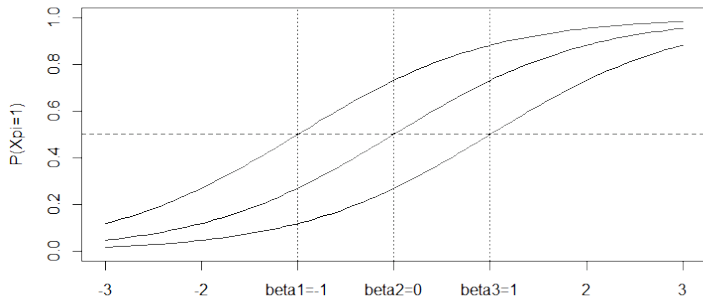
$$x_{pi} = \begin{cases} 1, & \text{wenn Antwort richtig/Zustimmung*} \\ 0, & \text{wenn Antwort falsch/Ablehnung*} \end{cases}$$

* von Person p auf Item i.

► also Matrix mit Nullen und Einsen

IRT-Bsp.: Das Rasch-Modell

- ▶ $P(X_{pi} = 1) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}, p = 1, \dots, P; i = 1, \dots, I$
 - ▶ θ_p wird bezeichnet als Personenparameter der Person p
 - ▶ β_i wird bezeichnet als Itemparameter des Items i
- ▶ Es ergeben sich also $P + I - 1$ Parameter (Restriktion nötig)
- ▶ Beispiel für drei Item-Response Funktionen



Eigenschaften des Rasch-Modells

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Gürer

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Suffiziente Statistiken sowohl für Item- als auch Personenparameter vorhanden (Randsummen der Antwortmatrix).
- ▶ Schätzung β über jML, cML oder mML
- ▶ θ ist herauskonditionierbar $\rightarrow$ cML-Schätzung der Itemparameter. Unabhängig von nuisance und konsistent.
- ▶ Jedes Item zeigt gleiche Diskrimination, dadurch Rangreihe der Items für jede Person identisch.
- ▶ Beliebtes Modell, da Bestimmung von θ unabhängig von konkreten Items (also β) immer gleich. Dadurch ist adaptives Testen ermöglicht, also Testung mit individueller Itemzusammenstellung.

- Zusätzliche Einführung eines Diskriminationsparameters α_i für jedes Item i:

$$P(X_{pi} = 1) = \frac{\exp(\alpha_i(\theta_p - \beta_i))}{1 + \exp(\alpha_i(\theta_p - \beta_i))}$$

Eigenschaften des 2PL-Modells

- ▶ Flexibler als das Rasch-Modell.
- ▶ Viele Eigenschaften fallen jedoch im Gegensatz zum Rasch-Modell weg.
 - ▶ Suffiziente Statistiken nun $\sum_{i=1}^I \alpha_i x_{pi}$. $\rightarrow$ Beim 2PL-Modell also ausschlaggebend, welche Items genau gelöst wurden.
 - ▶ Rangreihe der Items kann für Personen verschieden ausfallen.
 - ▶ Größere Stichproben zur zuverlässigen Schätzung der α_i notwendig.
- ▶ Spezifiziert man die α_i vor der Analyse, so zeigt das resultierende OPLM (One-Parameter-Logistic-Model, nicht zu verwechseln mit 1PL) wieder die wünschenswerten Eigenschaften des Rasch-Modells (Personenwerte wieder unabhängig von ItemSP identisch schätzbar).

- ▶ Weitere Einführung eines Rateparameters γ_i für jedes Item i :

$$P(X_{pi} = 1) = \gamma_i + (1 - \gamma_i) \frac{\exp(\alpha_i(\theta_p - \beta_i))}{1 + \exp(\alpha_i(\theta_p - \beta_i))}$$

Erweiterungen der IRT-Modelle

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Gürer

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ mehrdimensionale Modelle mit $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \dots, \theta_m)$.
 - ▶ mehrdimensionales Rasch-Modell
 - ▶ multidimensional random coefficient multinomial logit model (MRCMLM)
 - ▶ ...
- ▶ Erweiterungen auf mehrkategoriale Items, z.B. $x_{pi} \in \{0, 1, 2, 3, 4\}$.
 - ▶ (Generalised) Partial-Credit-Modell (GPCM/PCM)
 - ▶ Ratingscale-Modell (RSM)
 - ▶ ...
- ▶ Mischverteilungsmodelle
 - ▶ Mixed-Rasch-Modell
 - ▶ Hybrid Modelle
 - ▶ ...

CDM: Cognitive Diagnosis Models

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Güler

Übersicht

KTT

IRT

Datenformat

Rasch-Modell

2PL-Modell

3PL-Modell

Andere Erweiterungen

CDM

DINA-Modell

Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Grundidee: diskrete Skills als Grundlage, Diagnostik der Skills als Ziel.
- ▶ Skillvektor einer Person p $\alpha_p \in \{0, 1\}^K$ mit $K =$ Anzahl verschiedener Skills
- ▶ Bsp. Mathematikfähigkeiten
 - ▶ Skill 1: Addition
 - ▶ Skill 2: Multiplikation
 - ▶ Skill 3: Wurzelziehen
 - ▶ Skill 4: Punkt vor Strich
- ▶ Aufgabe: $2 + 2 \cdot 3 = ?$
 - ▶ benötigte Skills: 1,2,4 $\rightarrow$ Q-Matrix Eintrag $\mathbf{q}_i = (1, 1, 0, 1)$.
 - ▶ Die Q-Matrix enthält die Anforderungsprofile aller Items i des Tests.
- ▶ Verrechnung von α_p mit $\mathbf{q}_i$ zu $\eta_{pi} = \sum_{k=1}^K \alpha_{pk}^{q_{ik}}$

CDM-Bsp.: Das DINA-Modell

- ▶ DINA-Modell als Beispiel (Deterministic inputs, noisy "and" gate)
- ▶ Einfachstes CDM.

$$P_i(\alpha_p) = P(X_{pi} = 1 | \alpha_p) = g_i^{1-\eta_{pi}} (1 - s_i)^{\eta_{pi}}$$

- ▶ Interpretation:
 - ▶ g_i Rateparameter. Kommt zum tragen, wenn $\eta_{pi} = 0$, wenn Person p mindestens einen für Item i benötigten Skill k nicht beherrscht.
 - ▶ s_i Parameter für "Leichtsinnfehler" (Slips). Das Komplement kommt zum tragen, wenn mindestens alle benötigten Skills von Person p beherrscht werden.
- ▶ Besonderheit: deterministischer Anteil im Modell durch η_{pi} .
- ▶ Schätzung über mML mithilfe EM-Algorithmus.

- ▶ mehrstufige Skills (GDINA)
- ▶ weitere Parameter, wie z.B. Vollständigkeitsparameter (sind alle erforderlichen Skills in der Q-Matrix enthalten?)
- ▶ Kombination mit IRT-Anteilen, RUM (Reparametrized Unified Model)
- ▶ RUM mit Prioris für Skills: Fusion Model.

Weitere Begriffe und Ansätze psychometrischer Modelle

Der Versuch eines
Überblicks über
psychometrische
Modelle

Can Gürer

Übersicht

KTT

IRT

Datenformat
Rasch-Modell
2PL-Modell
3PL-Modell
Andere Erweiterungen

CDM

DINA-Modell
Erweiterungen

Weitere Begriffe
und Ansätze

References

- ▶ Forced-Choice Antwortformate, Präferenzantworten
- ▶ unfolding Antwortmodelle (statt dominance response)
 - ▶ Dominance-Response (alle heute behandelten Modelle):
monoton steigende Item-Response Funktionen
 - ▶ Unfolding-Modelle: eingipflige IRF, die
Zustimmungswahrscheinlichkeit sinkt nach beiden
Seiten ab (Näherelation).

Herzlichen Dank für die Aufmerksamkeit

Rost J (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Aus dem Programm Huber: Psychologie-Lehrbuch. Huber Hans.

von Davier M, Carstensen C (2007). *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*. Statistics for Social Science and Behavioral Sciences. Springer Science + Business Media, LLC, New York.