

## Grundlagen der Item-Response-Theorie

- Im Unterschied zur KTT werden in der IRT ausdrücklich latente, nicht direkt beobachtbare Persönlichkeitsvariablen angenommen, d.h. Merkmal und Verhalten im Test werden nicht gleichgesetzt („Modelle mit latenten Variablen“).
- Die IRT formuliert Modelle zu den Zusammenhängen zwischen den latenten Persönlichkeitsvariablen und den Antworten in psychologischen Tests.

---

---

---

---

---

---

---

## Grundlagen der Item-Response-Theorie

- Es werden Annahmen über die Zusammenhänge zwischen der individuellen Merkmalsausprägung und der Wahrscheinlichkeit\* für das Auftreten bestimmter Antworten\*\* formuliert.

\* → Daher auch probabilistische Testtheorie

\*\* → Daher Item-Response-Theorie

---

---

---

---

---

---

---

## Grundlagen der Item-Response-Theorie

### Vorteile gegenüber der KTT

- Die Schätzung der Merkmalsausprägungen (Personenparameter  $\theta_i$ ) und der Aufgabenmerkmale (Itemparameter) ist unabhängig von der Stichprobe und den verwendeten Aufgaben möglich.
- Die IRT erlaubt eine kriterienorientierte Interpretation der gemessenen Merkmalsausprägungen.

---

---

---

---

---

---

---

## Latent-Trait-Modelle

- Vorstellungen über den Zusammenhang zwischen *kontinuierlichen* latenten Merkmalen („latent trait“) und Manifestation werden in *Latent-Trait-Modellen* ausgedrückt.
- Für die manifesten Verhaltensvariablen werden in Abhängigkeit von der Ausprägungen des latenten Merkmals bestimmte Werte erwartet,
- Unterschiedliche Ausprägungen einer latenten Variablen bewirken *systematische* Unterschiede im beobachtbaren Verhalten.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

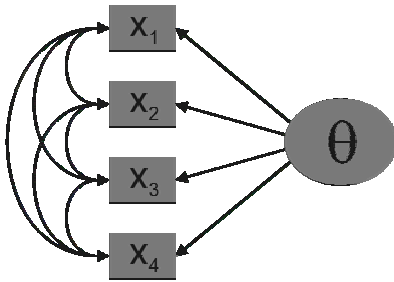
---

---

---

---

## Latente Variable als Quelle für Zusammenhänge zwischen den Items



Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

## Lokale stochastische Unabhängigkeit

- Zusammenhänge zwischen den manifesten Variablen (Items) kommen nur durch die Ausprägung des latenten Merkmals zustande.
- Wenn alle Items homogen sind, d.h. ein und dieselbe latente Dimension erfassen, so sollten ihre Korrelationen untereinander nur durch die Unterschiede in der latenten Dimension hervorgerufen werden.
- Die lokale stochastische Unabhängigkeit besteht darin, dass die Korrelation zwischen den Items auf einer lokalen Stufe der latenten Dimension verschwindet.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

## Itemcharakteristische Funktion

- Der Zusammenhang zwischen der Wahrscheinlichkeit für eine bestimmte Antwort und dem latenten Merkmal  $\theta$  wird durch die Itemcharakteristische Funktion (IC-Funktion) beschrieben.

---

---

---

---

---

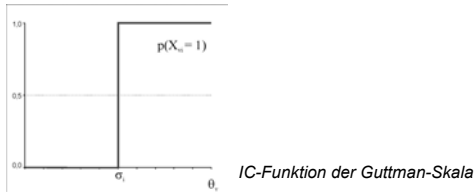
---

---

## Itemcharakteristische Funktion

- Deterministische IC-Funktion: Guttman-Skala

Im Guttman-Modell existieren nur Antwort- bzw. Lösungswahrscheinlichkeiten von 0 und 1, d.h. das Antwortverhalten kann mit Personenfähigkeit  $\theta_v$  und Itemschwierigkeit  $\sigma_i$  „sicher“ vorhergesagt werden.



---

---

---

---

---

---

---

## Deterministische vs. probabilistische IC-Funktionen

- Das deterministische Guttman-Modell ist für empirische Daten i.d.R. viel zu streng.
- Die meisten IRT-Modelle arbeiten mit Antwortwahrscheinlichkeiten zwischen null und eins.
- Diese IC-Funktionen sind dann i.e.S. *probabilistisch*.
- Das einfachste und verbreitetste probabilistische Testmodell ist das dichotome Raschmodell.

---

---

---

---

---

---

---

## Das dichotome Raschmodell

- Das dichotome Raschmodell für zweistufige Antworten kann z.B. auf Fragebogenitems mit einem „ja“/„nein“-Antwortformat oder Leistungsaufgaben mit den Bewertungsstufen „gelöst“/ „nicht gelöst“ angewendet werden.
- Es enthält folgende Parameter:
  - Jede Person hat eine individuelle latente Merkmalsausprägung  $\theta_v$ ,
  - Jedes Item hat eine Schwierigkeit  $\sigma_i$ .
- Personen- und Itemparameter werden auf derselben, eindimensionalen Skala dargestellt.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

---

## Das dichotome Raschmodell

- Das dichotome Raschmodell nimmt eine logistische IC-Funktion zwischen der individuellen Merkmalsausprägung  $\theta_v$  und der Lösungswahrscheinlichkeit eines Items  $i$  in Abhängigkeit von dessen Schwierigkeit  $\sigma_i$  an:

$$p(X_{vi} = 1) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}$$

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

---

## Das dichotome Raschmodell

Entscheidend für die Lösungswahrscheinlichkeit ist die Differenz zwischen der individuellen Merkmalsausprägung  $\theta_v$  und der Schwierigkeit  $\sigma_i$  des jeweiligen Items  $i$ .

$$p(X_{vi} = 1) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}$$

$p(X_{vi}=1)$ : Wahrscheinlichkeit, das ein Item  $i$  von Person  $v$  gelöst bzw. im Sinne einer hohen Merkmalsausprägung beantwortet wird,

$\theta_v$ : Merkmalsausprägung der Person  $v$ ,

$\sigma_i$ : Schwierigkeit des Items  $i$ .

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

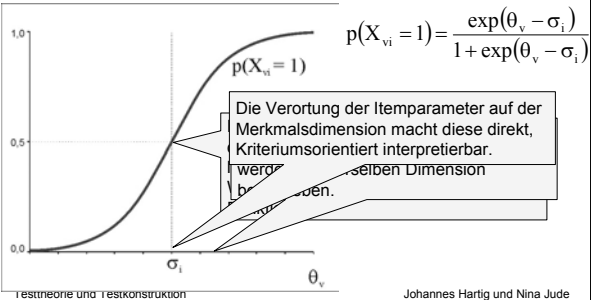
---

---

---

## Das dichotome Raschmodell

Itemcharakteristische Funktion des dichotomen Raschmodells




---

---

---

---

---

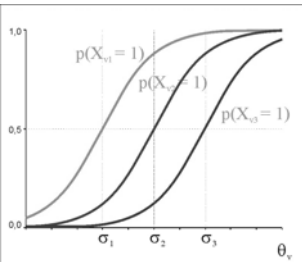
---

---

---

## Das dichotome Raschmodell

IC-Funktionen dreier rasch-homogener Items:  
Parallele Itemfunktionen




---

---

---

---

---

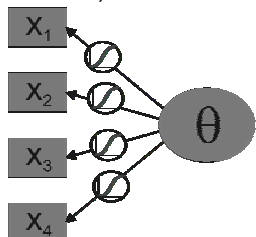
---

---

---

## Das dichotome Raschmodell

- Die IC-Funktion beschreibt den angenommenen Zusammenhang zwischen latenter und manifesten Variablen („link function“)




---

---

---

---

---

---

---

---

## Parameterschätzung im Raschmodell

- Im Rasch-Modell erfolgt die Schätzung der Personen- und Aufgabenparameter nach dem *Maximum-Likelihood-Prinzip*:
- Dabei beschreibt die Likelihoodfunktion (LF) die Wahrscheinlichkeit der Daten, unter der Annahme, dass das Modell gilt.
- Der Wert der LF gibt also an, wie wahrscheinlich die beobachteten Daten sind, wenn das Modell wirklich gilt.

---

---

---

---

---

---

---

## Parameterschätzung im Raschmodell

- Vereinfacht ausgedrückt werden sowohl die Itemparameter  $\theta_v$  als auch die Personenparameter  $\sigma_i$  gemeinsam so geschätzt, dass die Likelihood für die beobachteten Daten möglichst groß wird → *Maximum Likelihood*.
- Dieses Vorgehen erfolgt iterativ (schrittweise), d.h. die optimalen Werte der Parameter werden nicht durch das Lösen eindeutiger Gleichungssysteme ermittelt, sondern durch „ausprobieren“.
- Hierzu sind spezielle Computerprogramme notwendig.

---

---

---

---

---

---

---

## Parameterschätzung im Raschmodell

- Die Eigenschaften des Rasch-Modells erlauben eine Schätzung Parameter allein auf Basis der Zeilen- und Spaltensummen der Rohdatenmatrix („**suffiziente Statistiken**“). Mit anderen Worten:
- Die Schwierigkeit eines Items ist nur davon abhängig, von *wie vielen* Personen es gelöst wurde – nicht von *welchen* Personen;
- Die Fähigkeit einer Person ist nur davon abhängig, *wie viele* Items sie gelöst hat – nicht *welche* Items.

---

---

---

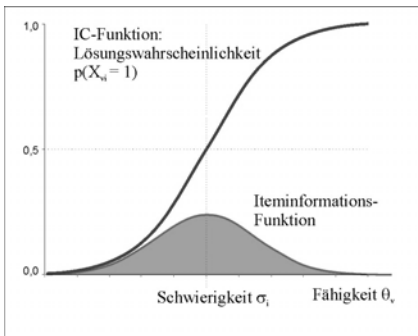
---

---

---

---

## Dichotomes Raschmodell: IC-Funktion und Iteminformations-Funktion



Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

---

## IRT-basiertes adaptives Testen

- Beim adaptiven Testen sollen die zu bearbeitenden Items der Fähigkeit der getesteten Person angepasst werden (auch „tailored testing“ = „maßgeschneidertes Testen“).
- Dies wird erreicht, indem Aufgaben gewählt werden, deren Schwierigkeit möglichst Nahe der Fähigkeit der Person liegt (Lösungswahrscheinlichkeit = 50%) und damit maximale Information liefern.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

---

## IRT-basiertes adaptives Testen

1. Wähle eine Aufgabe mit maximaler Information (d.h. deren Schwierigkeit der Fähigkeit entspricht)
2. Aufgabenvorgabe und -beantwortung
3. Schätze Personenfähigkeit
4. Schätze Messgenauigkeit
5. Messgenauigkeit zufrieden stellend oder maximale Itemzahl erreicht? Wenn NEIN, zurück zu (1).
6. Aktuelle Personenfähigkeit = Testergebnis

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

---

## IRT-basiertes adaptives Testen

- Beim adaptiven Testen bearbeiten verschiedene Personen unterschiedliche Items
- Ohne IRT-Skalierung wäre es nicht möglich, für die verschiedenen Personen vergleichbare Testwerte zu ermitteln.
- Voraussetzung für IRT-basiertes adaptives Testen ist idealerweise in großer Pool erprobter Items mit bekannten Schwierigkeiten.
- Adaptive Tests sind typischerweise Computerbasiert.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

## Das ordinale Raschmodell

- Das ordinale Raschmodell beschreibt die Antwortwahrscheinlichkeiten für Items mit mehrstufigen Antworten, z.B. Ratingskalen in Persönlichkeitsfragebögen oder verschiedene Punktezahlen bei der Bewertung von Leistungstestaufgaben.
- Bei Leistungstests kann z.B. für nicht ganz richtig gelöste Aufgaben ein Teil der maximal möglichen Punkte gegeben werden, daher wird dieses Modell auch als „Partial Credit-Modell“ bezeichnet.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---

## Das ordinale Raschmodell

- Im ordinalen Raschmodell wird - entsprechend den Items im dichotomen Raschmodell - für *jede Stufe eines Items* eine logistische Funktion angenommen, die die Wahrscheinlichkeit beschreibt, dass eine Person mit einer Merkmalsausprägung  $\theta_v$  diese Stufe erreicht oder überschreitet.

Testtheorie und Testkonstruktion

Johannes Hartig und Nina Jude

---

---

---

---

---

---

---



## Das ordinale Raschmodell

- Die Schwierigkeit der Itemstufen  $k$  eines Items  $i$  werden durch Schwellenparameter  $\tau_{ik}$  beschrieben.
- Die Funktionen für die einzelnen Schwellen eines ordinalen Items entsprechen jeweils der IC-Funktion eines einzelnen Items im dichotomen Raschmodell:

$$p(X_{vik} \geq k) = \frac{\exp(\theta_v - \tau_{ik})}{1 + \exp(\theta_v - \tau_{ik})}$$

---

---

---

---

---

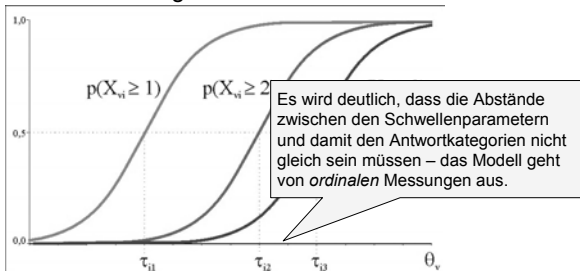
---

---

---

## Das ordinale Raschmodell

Beispiel. Schwellenfunktionen für **ein Item** mit 4 ordinalen Kategorien 0 bis 3.




---

---

---

---

---

---

---

---

## Das ordinale Raschmodell

- Kategorienwahrscheinlichkeiten im ordinalen Raschmodell: Wahrscheinlichkeit für das Auftreten einer Antwort in Kategorie  $k$  in Item  $i$
- Es werden kumulierte Schwierigkeiten  $\sigma_{ik}$  für jede Schwelle gebildet, die sich aus dem Schwellenparameter  $\tau_{ik}$  der Schwelle  $k$  und aller vorhergehenden Schwellen ergeben:

$$\sigma_{ik} = \sum_{s=1}^k \tau_{is}$$

---

---

---

---

---

---

---

---

## Das ordinale Raschmodell

- Mit Hilfe der kumulierten Schwellenschwierigkeiten  $\sigma_{ik}$  kann die Wahrscheinlichkeit für das Auftreten einer Antwort in Kategorie k beschrieben werden:

$$p(X_{vik} = k) = \frac{\exp(k \cdot \theta_v - \sigma_{ik})}{\sum_{s=0}^m \exp(s \cdot \theta_v - \sigma_{is})} \quad \sigma_{ik} = \sum_{s=1}^k \tau_{is}, \quad \sigma_{i0} = 0$$

(vgl. Rost, 2003)

---

---

---

---

---

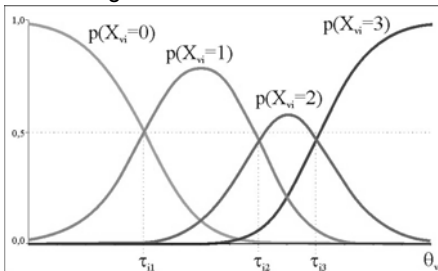
---

---

---

## Das ordinale Raschmodell

- Wahrscheinlichkeit für das Auftreten einer Antwort in Kategorie k in Item i



---

---

---

---

---

---

---

---

## Latent-Class-Modelle

- Latent-Trait-Modelle wie das Raschmodell gehen von quantitativen latenten Variablen aus.
- Latent-Class-Modelle nehmen qualitative latente Variablen zur Charakterisierung von Personenunterschieden an.
- Lösungswahrscheinlichkeiten jedes einzelnen Items unterscheiden sich zwischen den latenten Klassen (Personengruppen).

---

---

---

---

---

---

---

---

## Latent-Class-Modelle

- Die Klassen sind durch spezifische Antwortmuster charakterisiert; innerhalb der Klassen wird stochastische Unabhängigkeit angenommen.
- Die Anzahl der anzunehmenden Klassen muss bei der Analyse vorgegeben werden die Klassengrößen werden geschätzt.
- Die Klassenzuordnung von Personen erfolgt probabilistisch auf Basis der Ähnlichkeiten der beobachteten mit den klassenspezifischen Antwortmustern.

---

---

---

---

---

---

---