

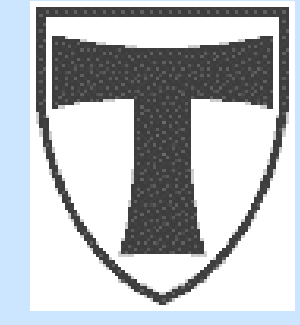
# Beurteilung der Validität von Wissensstrukturen durch den Diskrepanzindex

Florian Wickelmaier<sup>1</sup> und Jürgen Heller<sup>2</sup>



<sup>1</sup>Department of Acoustics, Aalborg University, 9220 Aalborg, Dänemark

<sup>2</sup>Abteilung Psychologische Methodik, Justus-Liebig-Universität Gießen, 35394 Gießen, Deutschland



## Theorie der Wissensstrukturen

Die Theorie der Wissensstrukturen (Doignon & Falgagne, 1985) bietet einen formalen Rahmen zur Erfassung und effizienten Diagnose des Wissens von Probanden in einem bestimmten Wissensbereich:

- Ein Wissensbereich wird als Menge  $Q$  von Aufgaben formalisiert.
- Menge  $K \subseteq Q$  aller Aufgaben, die ein Proband lösen kann, kennzeichnet dessen Wissenszustand.
- Eine Wissensstruktur beschreibt die Menge  $\mathcal{K}$  der möglichen Wissenszustände.

Aufgrund von Flüchtigkeitsfehlern und/oder Raten wird sich im Allgemeinen das Antwortmuster  $R \subseteq Q$  eines Probanden von seinem Wissenszustand  $K$  unterscheiden.

### Problem: Empirische Validierung einer Wissensstruktur

Wie läßt sich entscheiden, ob eine gegebene Wissensstruktur die empirisch erhobenen Antwortmuster adäquat beschreibt?

## Diskrepanzindex

Als Beschreibung der Ähnlichkeit einer Wissensstruktur  $\mathcal{K}$  zur Menge  $\mathcal{R}$  der Antwortmuster wurde der Diskrepanzindex vorgeschlagen (Kambouri et al., 1994):

- Die Ähnlichkeit des Antwortmusters  $R$  zum Wissenszustand  $K$  ergibt sich durch die symmetrische Mengendifferenz

$$d(R, K) = |(R \setminus K) \cup (K \setminus R)|$$

(Anzahl der  $R$  und  $K$  unterscheidenden Aufgaben).

- Zu jedem Antwortmuster  $R$  bestimmt man die minimale Distanz zu einem der Wissenszustände in der Wissensstruktur  $\mathcal{K}$  (bezeichnet mit  $d(R, \mathcal{K})$ ) und betrachtet die sogenannte Diskrepanzfunktion, d. h. die Verteilungsfunktion der minimalen Distanzen über alle Antwortmuster.
- Der Erwartungswert der Diskrepanzfunktion heißt Diskrepanzindex.

### Der Diskrepanzindex als Validitätsmaß

Der Diskrepanzindex gibt an, wie stark sich die Antwortmuster im Durchschnitt von den Wissenszuständen unterscheiden: Je größer der Wert, desto schlechter die Anpassung der Wissensstruktur an die Daten. Es stellt sich folgendes Problem:

Ab welchem Wert kann die Wissensstruktur nicht mehr als adäquate Beschreibung der Antwortmuster angesehen werden?

## Probabilistische Wissensstrukturen und Diskrepanzindex

Obwohl dem Diskrepanzindex ein rein deterministischer Zusammenhang zwischen der Wissensstruktur und den Antwortmustern zugrundeliegt, lassen sich damit Schätzfunktionen gewinnen für

- die Auftretenswahrscheinlichkeit  $p(K)$  für jeden Wissenszustand  $K$ :

$$\hat{p}(K) = \frac{1}{N} \sum_{\substack{R \in \mathcal{R}: \\ d(R, K) = d(R, \mathcal{K})}} N(R)/\alpha(R), \quad (1)$$

- die Fehlerwahrscheinlichkeit  $\beta_q$ , d. h. die bedingte Wahrscheinlichkeit für eine falsche Antwort bei Aufgabe  $q$ , obwohl der Proband die Aufgabe  $q$  eigentlich beherrscht:

$$\hat{\beta}_q = \frac{\sum_{K \in \mathcal{K}: q \in K} \sum_{\substack{R \in \mathcal{R}: q \notin R \\ d(R, K) = d(R, \mathcal{K})}} N(R)/\alpha(R)}{\sum_{K \in \mathcal{K}: q \in K} \sum_{\substack{R \in \mathcal{R}: \\ d(R, K) = d(R, \mathcal{K})}} N(R)/\alpha(R)}, \quad (2)$$

- die Ratewahrscheinlichkeit  $\eta_q$ , d. h. die bedingte Wahrscheinlichkeit für eine richtige Antwort bei Aufgabe  $q$ , obwohl der Proband die Aufgabe  $q$  eigentlich nicht beherrscht:

$$\hat{\eta}_q = \frac{\sum_{K \in \mathcal{K}: q \notin K} \sum_{\substack{R \in \mathcal{R}: q \in R \\ d(R, K) = d(R, \mathcal{K})}} N(R)/\alpha(R)}{\sum_{K \in \mathcal{K}: q \notin K} \sum_{\substack{R \in \mathcal{R}: \\ d(R, K) = d(R, \mathcal{K})}} N(R)/\alpha(R)} \quad (3)$$

( $\alpha(R)$  gibt an, zu wievielen Wissenszuständen das Antwortmuster  $R$  minimale Distanz aufweist,  $N(R)$  bezeichnet die absolute Häufigkeit des Antwortmusters  $R$ ).

## Grundidee der Validitätsprüfung

Beurteilung der empirischen Validität einer Wissensstruktur  $\mathcal{K}$  im Rahmen eines statistischen Tests:

- Nullhypothese  
Die Antwortmuster gehen aus der Wissensstruktur  $\mathcal{K}$  hervor, mit unbekanntem stochastischen Parametern  $p(K)$ ,  $\beta_q$  und  $\eta_q$ .
- Teststatistik  
Ermittlung des Diskrepanzindex zwischen den empirisch erhobenen Antwortmustern und der Wissensstruktur  $\mathcal{K}$ .
- Prüfverteilung  
– Schätzung der stochastischen Parameter der probabilistischen Wissensstruktur  $\mathcal{K}$   
– Monte-Carlo-Simulation von Antwortmustern auf der Basis der probabilistischen Wissensstruktur  $\mathcal{K}$  und Berechnung des resultierenden Diskrepanzindex zwischen simulierten Antwortmustern und Wissensstruktur  $\mathcal{K}$   
⇒ Die Verteilung der Diskrepanzindizes dient als Prüfverteilung.
- Testentscheidung  
Ablehnung der Nullhypothese, falls der empirisch gefundene Diskrepanzindex größer ist als das 95%-Perzentil der Prüfverteilung.

## Empirische Anwendung

### Antwortmuster

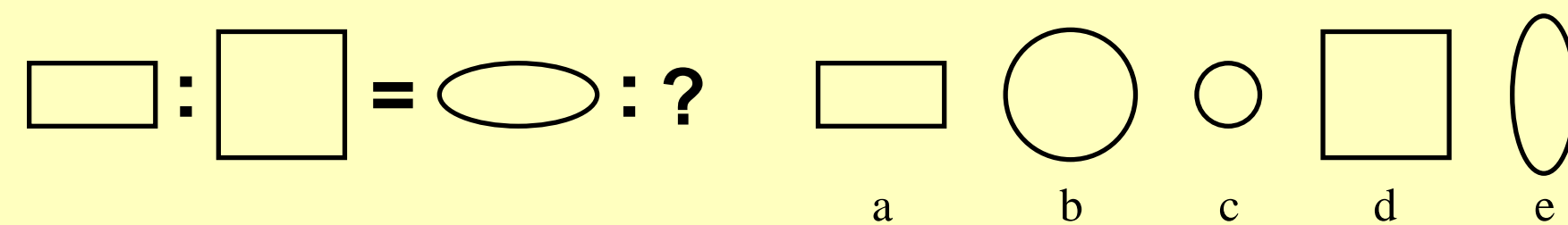
Reanalyse der Daten aus einer an der Karl-Franzens-Universität Graz durchgeführten Untersuchung zum induktiven Denken (Wesiak & Albert, 2001). Der verwendete Test bestand aus vier Subtests mit je fünf Aufgaben: verbale Analogien, figurale Analogien, verbale Reihenergänzung und figurale Reihenergänzung.

### Beispielaufgaben

Aufgabe a (verbale Analogien):

Herz zu Pumpe wie Gehirn zu ?  
a) denken b) Zentrale c) Verstand d) Kopf e) Nerven.

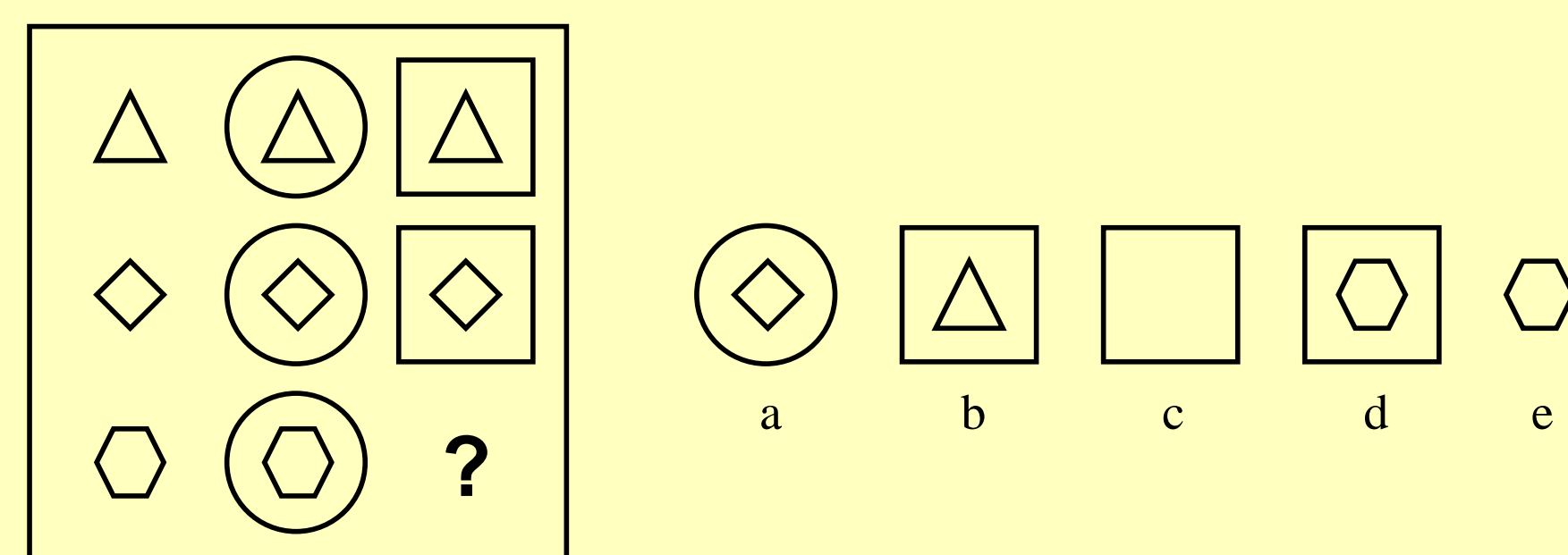
Aufgabe h (figurale Analogien):



Aufgabe k (verbale Reihenergänzung):

25 23 20 18 15 ?  
a) 12 b) 17 c) 13 d) 14 e) 11.

Aufgabe q (figurale Reihenergänzung):

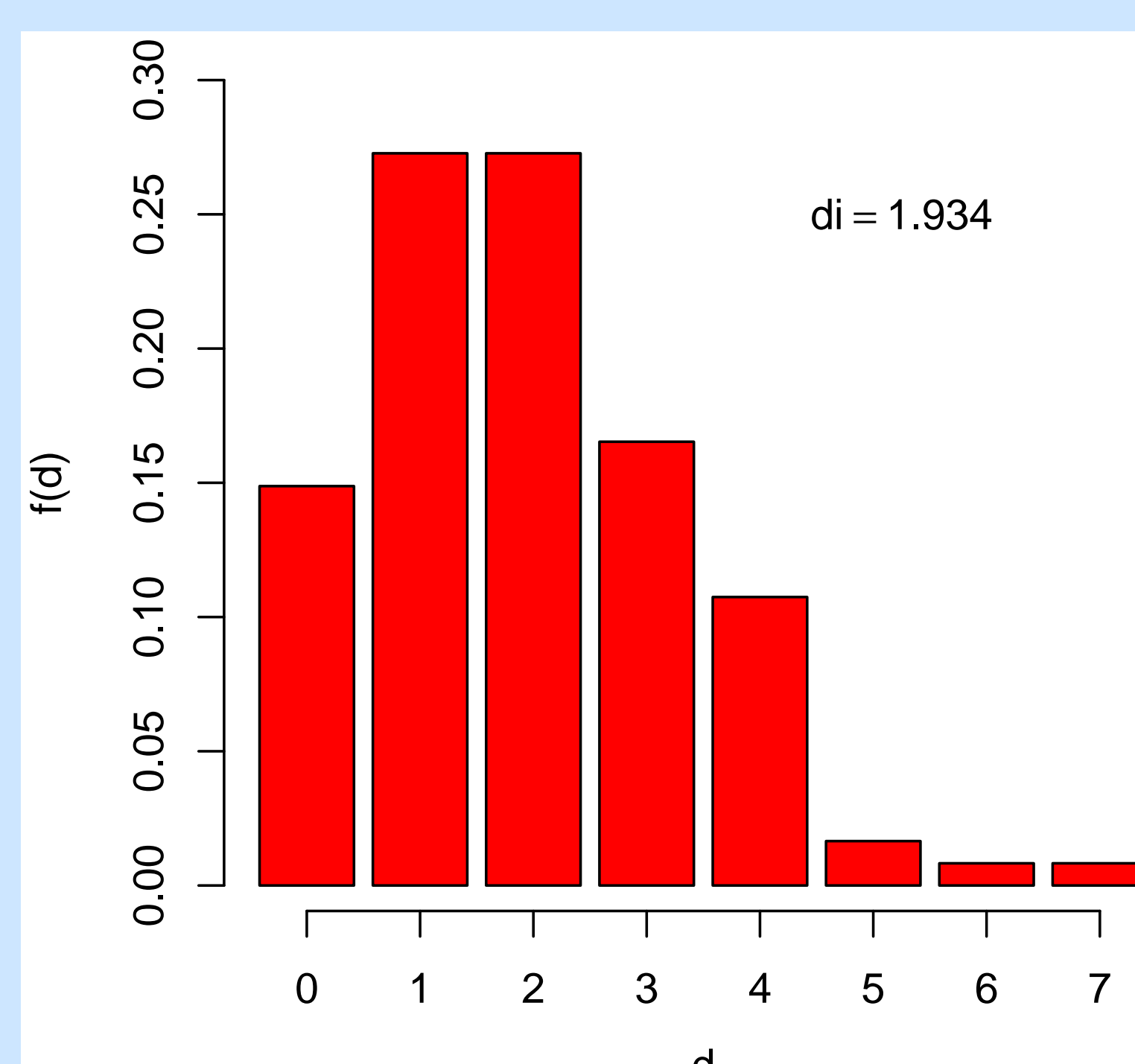


## Wissensstruktur

Zur Beschreibung der Daten von 121 Probanden schlugen Wesiak & Albert (2001) eine Wissensstruktur  $\mathcal{K}$  vor, die aus 438 Wissenszuständen bestand (bei insgesamt  $2^{20}$  möglichen Wissenszuständen!). Die empirische Validität dieser durch Item-Tree-Analyse (Leeuwe, 1974) ermittelten Wissensstruktur  $\mathcal{K}$  wird nachfolgend untersucht.

## Ergebnisse

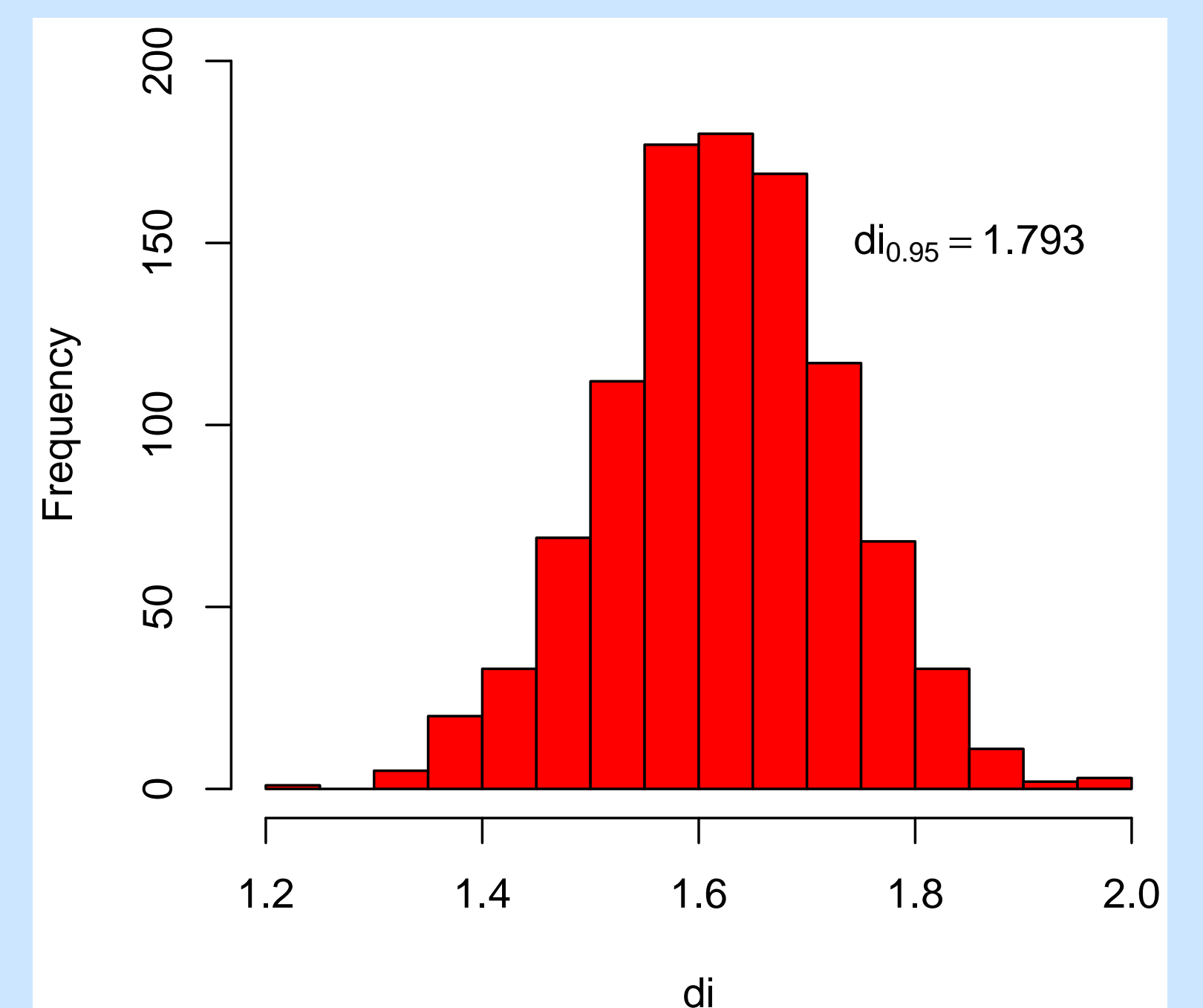
Die folgende Abbildung zeigt die Verteilung der minimalen Distanzen. Der Diskrepanzindex  $di = 1.934$  zwischen den Antwortmustern und der unterlegten Wissensstruktur ist der Erwartungswert dieser Verteilung.



Die gemäß (2) und (3) geschätzten Fehler- und Ratewahrscheinlichkeiten für alle 20 Aufgaben ( $a-t$ ):

	Parameter	Schätzung	Parameter	Schätzung
Fehlerwahrscheinlichkeit	$\beta_a$	0.0965	$\beta_k$	0.0622
	$\beta_b$	0.0309	$\beta_l$	0.1197
	$\beta_c$	0.0087	$\beta_m$	0.1983
	$\beta_d$	0.0525	$\beta_n$	0.1512
	$\beta_e$	0.0485	$\beta_o$	0.0331
	$\beta_f$	0.0158	$\beta_p$	0.0606
	$\beta_g$	0.2003	$\beta_q$	0.0000
	$\beta_h$	0.0996	$\beta_r$	0.1203
	$\beta_i$	0.0472	$\beta_s$	0.1068
	$\beta_j$	0.0238	$\beta_t$	0.0000
Ratewahrscheinlichkeit	$\eta_a$	0.0634	$\eta_k$	0.0000
	$\eta_b$	0.0162	$\eta_l$	0.0733
	$\eta_c$	0.1267	$\eta_m$	0.0000
	$\eta_d$	0.1863	$\eta_n$	0.0000
	$\eta_e$	0.2102	$\eta_o$	0.1376
	$\eta_f$	0.0295	$\eta_p$	0.0000
	$\eta_g$	0.0000	$\eta_q$	1.0000
	$\eta_h$	1.0000	$\eta_r$	0.0000
	$\eta_i$	0.0482	$\eta_s$	0.1938
	$\eta_j$	0.0000	$\eta_t$	0.1324

Mit diesen Parameterschätzungen für die unterliegende probabilistische Wissensstruktur  $\mathcal{K}$  werden im Rahmen einer Monte-Carlo-Simulation Diskrepanzindizes berechnet (1000 Werte). Es ergibt sich folgende Verteilung:



Als 95%-Perzentil der resultierenden Verteilung bestimmt man den Wert 1.793. Der empirisch gefundene Diskrepanzindex  $di = 1.934$  fällt somit in den Ablehnungsbereich.

### Schlußfolgerung: Empirische Validität der Wissensstruktur

Der statistische Test des Diskrepanzindex führt zur Ablehnung der Hypothese, dass die Wissensstruktur  $\mathcal{K}$  den empirisch erhobenen Antwortmustern zugrunde liegt.

## Diskussion

Der Diskrepanzindex ordnet jedem Antwortmuster den Wissenszustand mit geringster Distanz zu. Dieser unterscheidet sich aber möglicherweise stark vom wahren Wissenszustand des Probanden. Alle auf der Basis der minimalen Distanzen berechneten Parameterschätzer sind deshalb verzerrt. Insbesondere werden die Fehler- und Ratewahrscheinlichkeiten unterschätzt. Wegen der auftretenden Konfundierung der Validität der Wissensstruktur mit der Güte des Parameterschätzverfahrens erscheint der Diskrepanzindex – zumindest ohne weitere Modifikationen – als Grundlage eines statistischen Tests nur bedingt geeignet.

### Ausblick

Die geschätzten (minimalen) Fehler- und Ratewahrscheinlichkeiten liefern, auch abgesehen von statistischen Signifikanzüberlegungen, wertvolle Informationen über strukturelle Aspekte der den Antwortmustern zugrunde liegenden Wissensstruktur.

## Literatur

- Doignon, J.-P. & Falgagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175-196.
- Kambouri, M., Koppen, M., Villano, M. & Falgagne, J.-C. (1994). Knowledge assessment: tapping human expertise by the QUERY routine. *International Journal of Human-Computer Studies*, 40, 119-151.
- Leeuwe, J. F. J. v. (1974). Item tree analysis. *Nederlands Tijdschrift voor de Psychologie*, 29, 475-484.
- Wesiak, G. & Albert, D. (2001). Ordering Inductive Reasoning Tests based on Knowledge Space Theory. Poster presented at the 5th Meeting of the German Cognitive Science Society, Leipzig, Germany, September 2001.