

Andreas Mokros

# Testkonstruktion, Testen und Entscheiden

Fakultät für  
**Psychologie**

---

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Wir weisen darauf hin, dass die vorgenannten Verwertungsalternativen je nach Ausgestaltung der Nutzungsbedingungen bereits durch Einstellen in Cloud-Systeme verwirklicht sein können. Die FernUniversität bedient sich im Falle der Kenntnis von Urheberrechtsverletzungen sowohl zivil- als auch strafrechtlicher Instrumente, um ihre Rechte geltend zu machen.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m<sup>2</sup>, weiß), hergestellt aus 100 % Altpapier.

## Inhaltsverzeichnis

Inhaltsverzeichnis .....	3
Abbildungsverzeichnis .....	5
Tabellenverzeichnis.....	7
Einführung.....	8
Lehr- und Lernziele.....	9
1 Rückbezug auf die Klassische Testtheorie.....	10
1.1 Einleitung.....	10
1.2 Hintergrund .....	10
1.2.1 Axiome und Folgerungen.....	10
1.2.2 Reliabilitätsschätzung.....	12
1.3 Einschränkungen der Klassischen Testtheorie.....	13
2 Probabilistische Testmodelle .....	15
2.1 Einleitung.....	15
2.2 Übersicht über logistische Testmodelle für dichotome Itemantworten.....	15
2.2.1 Das 1PL-Modell (Rasch-Modell) .....	15
2.2.2 Das 2PL-Modell (Birnbaum-Modell) .....	15
2.2.3 Das 3PL-Modell (Modell mit Rateparameter).....	15
2.3 Das 1PLM (Rasch-Modell) im Detail .....	16
2.3.1 Herleitung.....	16
2.3.2 Schätzung der Itemparameter und spezifische Objektivität.....	21
2.3.3 Personenparameter und deren Genauigkeit.....	23
2.3.3.1 Die Schätzung des Fähigkeitsparameters $\theta$ .....	23
2.3.3.2 Die Informationsfunktion .....	26
2.3.4 Vorläufer.....	28
2.3.4.1 Die deterministische Guttman-Skalierung.....	28
2.3.4.2 Das probabilistische Normal-Ogiven-Modell .....	30
2.4 Das 2PLM (Birnbaum-Modell).....	31
2.5 Das 3PLM.....	32
2.6 Andere probabilistische Testmodelle.....	33
2.6.1 Logistische Testmodelle für mehrkategoriale Itemantworten .....	33
2.6.2 Das <i>One-Parameter-Logistic Model</i> .....	34
2.6.3 Der Begriff der <i>Item Response Theory</i> .....	35

2.6.4	Das nichtparametrische Modell nach Mokken.....	35
2.7	Adaptives Testen .....	37
2.7.1	Allgemeine Erläuterungen.....	37
2.7.2	Varianten .....	38
2.8	Beispiele.....	39
3	Klassifikatorische Modelle .....	40
3.1	Einleitung.....	40
3.2	Clusteranalyse .....	40
3.3	<i>Latent Class</i> -Analyse .....	41
3.3.1	Ein Beispiel .....	43
3.4	Multivariate Klassifikation .....	48
4	Diagnostische Indizes.....	49
4.1	Sensitivität und Spezifität.....	49
4.2	Güte des Auswahlresultates.....	51
4.2.1	Taylor-Russell-Tafeln .....	51
4.2.2	Inkrementeller Nutzen .....	52
4.2.2.1	Naylor-und-Shine-Modell .....	53
4.2.2.2	Geldwerter Entscheidungsnutzen .....	53
4.2.3	Grenzwertoptimierungskurven.....	54
4.2.3.1	Optimaler Schwellenwert .....	56
4.2.3.2	Kategorienspezifischer <i>Likelihood</i> -Quotient .....	58
4.2.3.3	Die Fläche unter einer Grenzwertoptimierungskurve .....	58
4.3	Diagnostische Indizes als Stichprobenschätzwerte.....	59
4.4	Diagnostische Irrtümer als Fehler 1. und 2. Art .....	60
5	Bayes'sches Schließen .....	61
5.1	Optimale Entscheidungen ohne Daten .....	61
5.1.1	Das Minimax-Prinzip .....	61
5.1.2	Das Bayes-Prinzip.....	62
5.2	Das Theorem von Bayes.....	63
6	Literatur .....	67
7	Anhang.....	72

## Abbildungsverzeichnis

Abb. 1: Unterschiedliche Messfehler bei unterschiedlichen Messwertausprägungen (nach Wottawa, 1980, S. 42) .....	13
Abb. 2: Transformation von Wahrscheinlichkeit (oben) in Wettquotienten (unten; nach Rost, 2004, S. 117).....	16
Abb. 3: Logarithmierung von Wettquotienten (nach Rost, 2004, S. 118).....	17
Abb. 4: Abbildungen 2 und 3 in einem Schritt – die Logit-Transformation (nach Rost, 2004, S. 118) .....	17
Abb. 5: Vergleich der Logit-Funktion (durchgezogen) mit einem linearen Verlauf (gestrichelt; mod. nach Rost, 2004, S. 118) .....	18
Abb. 6 Itemcharakteristische Funktion (ICC) für Item 27 aus dem KFT-Datensatz .....	22
Abb. 7 Itemcharakteristische Funktionen (ICCs) für die fünf Items aus dem KFT-Datensatz.....	23
Abb. 8 Testcharakteristische Funktion (TCC) für die KFT-Daten .....	26
Abb. 9 Informations- (durchgezogen) und itemcharakteristische Funktion (ICC, gestrichelt) für Item 19.....	26
Abb. 10 Testinformationsfunktion für die fünf KFT-Items.....	27
Abb. 11 Testcharakteristische Funktion (TCC) für die KFT-Daten (durchgezogen) mit 90%-Konfidenzintervall (gestrichelt).....	28
Abb. 12 Normal-Ogive (durchgezogene Linie) und ICC nach dem 1PL-(Rasch-)Modell für Item 27 der KFT-Daten (gestrichelte Linie).....	30
Abb. 13 Itemcharakteristische Funktionen (ICCs) für die fünf Items aus dem KFT-Datensatz nach dem 2PL-(Birnbaum-) Modell .....	31
Abb. 14 Itemcharakteristische Funktion (ICC) für ein Item nach dem 3PLM mit $\sigma_i = 2$ , $\beta_i = 1$ und $\gamma_i = .20$ .....	33
Abb. 15 Veranschaulichung der Schwellenwerte in drei Rasch-Modellen für mehrkategoriale Antwortformate, bezogen auf jeweils drei Items (nach: Rost, 2024, S. 234) .....	34
Abb. 16 Drei itemcharakteristische Funktionen gemäß dem Modell von Mokken.....	36
Abb. 17 Ergebnisse einer LCA mit drei Klassen für die KFT-Daten.....	43
Abb. 18 Bivariate Normalverteilung Standardnormalverteilung ( $r = .80$ ).....	48
Abb. 19 Sensitivität (Richtig-Positiv-Rate, RPR) und Spezifität (Richtig-Negativ-Rate, RNR) sowie deren Komplemente (Falsch-Negativ-Rate, FNR, bzw. Falsch-Positiv-Rate, FPR); fiktive Daten; Trennwert bei 17.5 Punkten. ....	50
Abb. 20 Selektionsschema (adaptiert nach Taylor & Russell, 1939, S. 567) .....	51
Abb. 21 Erfolgsrate (PPV) in Abhängigkeit von Selektionsrate und Validität, bei gegebener Basisrate von .50 (adaptiert nach Taylor & Russell, 1939, S. 569) .....	52

---

Abb. 22 Zuwachs im Kriterium in der ausgewählten Gruppe (in Standardabweichungseinheiten, $z$ ) in Abhängigkeit von der Selektionsrate .....	53
Abb. 23 Erstellung einer Grenzwertoptimierungs- (ROC-) Kurve: Koordinatenpunkte für die neun Kategorien aus dem Datenbeispiel .....	55
Abb. 24 Grenzwertoptimierungs- (ROC-) Kurve für die Koordinatenpunkte der neun Kategorien aus dem Datenbeispiel, Steigungstangenten (durchgezogene Linien) an den Punkten 3 und 5, Steigungstangente für den optimalen Trennwert (gestrichelte Linie) sowie Zufallserwartung (gepunktete Linie) .....	56
Abb. 25 Fläche unter der Grenzwertoptimierungskurve .....	58
Abb. 26 Veranschaulichung des Theorems von Bayes anhand dreier Krankheiten und eines gemeinsamen Symptoms .....	64

## Tabellenverzeichnis

Tab. 1.1: Annahmen der Testmodelle im Rahmen der Klassischen Testtheorie .....	11
Tab. 2.1: Schätzung der Itemparameter für die KFT-Daten in der Gesamt- und in Teilstichproben, mit Standardfehlern (in runden Klammern) und klassischen Schwierigkeitsindizes <sup>a)</sup> [in eckigen Klammern] .....	21
Tab. 2.2: Geschätzte Personenparameter <sup>a)</sup> für verschiedene Summenscores .....	24
Tab. 2.3: Geschätzte Personenparameter für die KFT-Daten nach der Expected A Posteriori (EAP)-Methode.....	25
Tab. 2.4: Geschätzte Lösungswahrscheinlichkeiten der KFT-Items in Abhängigkeit vom Summenscore gemäß unterschiedlicher Schätzmethoden .....	25
Tab 2.5: Lösungen von 6 Versuchspersonen (v) über 5 dichotome Items (i), verschieden geordnet .....	29
Tab. 2.6: Veranschaulichung von sogenannten Guttman-Fehlern .....	29
Tab. 3.1: Ergebnisse einer LCA mit drei Klassen für die KFT-Daten .....	43
Tab. 3.2: Mittlere Zuordnungswahrscheinlichkeiten .....	45
Tab. 3.3: Entscheidung über die Anzahl der zu extrahierenden Klassen .....	45
Tab. 3.4: Modellvergleiche: Unterschiede in den BIC-Werten, analoge Bayes-Faktoren (gerundet) und Interpretation .....	47
Tab. 4.1: Daten für die Erstellung einer Grenzwertoptimierungs-(ROC-) Kurve.....	54
Tab. 4.2: Vierfelderschema von Testergebnis und tatsächlichem Status .....	59
Tab. 4.3: Ausgewählte diagnostische Indizes mit 95%-Konfidenzintervallen.....	60
Tab. 5.1: Schadenstabelle für ein Beispiel ohne Daten .....	61
Tab. 5.2: Schadenstabelle für ein Beispiel ohne Daten mit Minimax-Optimalitätskriterium.....	62
Tab. 5.3: Schadenstabelle für ein Beispiel ohne Daten mit A-Priori-Wahrscheinlichkeiten ( $\pi$ ) und Bayes-Optimalitätskriterium .....	62

## Einführung

Der Kurs „Testkonstruktion, Testen und Entscheiden“ ist so angelegt, dass Sie Ihre grundlegenden Kenntnisse in der Entwicklung, Auswertung und Interpretation psychologischer Tests sowie in der Psychologischen Diagnostik erweitern.

Nach einer Wiederholung wesentlicher Aspekte der Klassischen Testtheorie liegt der Schwerpunkt auf der Probabilistischen Testtheorie. Neben dem Rasch-Modell und Modellen der Item-Response-Theorie (IRT) mit zwei bzw. drei Parametern werden unter anderem auch nicht-parametrische Varianten (Guttman- und Mokken-Skalierung) sowie Erweiterungen für *Rating*-Skalen vorgestellt. Die Modelle der Probabilistischen Testtheorie stellen die Voraussetzung für adaptives Testen dar.

Nach dieser Darstellung von geeigneten Methoden, um latente Eigenschaften zu messen, geht es um Klassifikation statt Quantifizierung: Nach einer kurzen Beschreibung der Cluster-Analyse wird die Latente Klassenanalyse als Gruppierungsverfahren beschrieben. Ferner wird kurz erläutert, wie Personen bestehenden Kategorien oder Gruppen zugeordnet werden können.

Nachfolgend geht es um diagnostische Kennwerte als Grundlagen für psychologische Auswahlentscheidungen oder Diagnosen (bzw. als Maße für deren Güte). Sie lernen kennen, wie man den bestmöglichen Trennwert einer Skala bestimmt, aber auch, wie die Effektivität von Auswahlentscheidungen angesichts der Tauglichkeit des Testverfahrens beurteilt werden kann. Damit bewegen wir uns bereits im Bereich des Entscheidens. Diesbezüglich lernen Sie zwei Methoden für Entscheidungsregeln ohne Daten kennen und schließlich – um sowohl Wahrscheinlichkeit als auch Validität in Entscheidungen mit einzubeziehen – die Grundlagen des Bayes'schen Schließens.

Das didaktische Konzept dieses Kurses beruht auf vier Säulen:

- 1) Der vorliegende Studienbrief soll Ihnen die wesentlichen Inhalte, Hintergründe und Herleitungen vermitteln.
- 2) Die Literaturempfehlungen am Ende jedes Kapitels liefern Ihnen eine tiefergehende und umfassendere Darstellung der jeweiligen Themen. Die Literatur wird in der Regel über die Bibliothek der FernUniversität in elektronischer Form zur Verfügung gestellt.
- 3) Die *Online*-Vorlesungen bieten Erläuterungen zu den Inhalten des Kurses, auch anhand von Beispielen und Veranschaulichungen.
- 4) In der *Moodle*-Umgebung zu diesem Kurs werden weitere Inhalte verfügbar gemacht; zudem ermöglicht sie Austausch und Diskussion.

Der Zugang zu *Moodle* befindet sich unter: <https://moodle.fernuni-hagen.de>

Ich wünsche Ihnen viel Spaß (und möglichst viele Aha-Erlebnisse) bei der Beschäftigung mit diesem Kurs sowie viel Erfolg,

Ihr

Andreas Mokros