# 1

# Bedeutung des Rasch-Modells für die Entwicklung psychologischer Tests

In vielen Situationen im schulischen, beruflichen und klinischen Alltag werden psychologische Tests verwendet, z.B. um die mathematische Kompetenz von Schülern, die Eignung von Bewerbern, die Kundenzufriedenheit, bestimmte Persönlichkeitseigenschaften oder die Neigung zu Depressionen zu messen.

Bei den zu messenden Eigenschaften der Personen handelt es sich dabei um latente, d.h. nicht direkt beobachtbare, Merkmale. Um diese Merkmale zu erfassen, werden den Personen in einem psychologischen Test mehrere Aufgaben oder Fragen gestellt, deren Beantwortung Aufschluss über die interessierende Eigenschaft geben soll.

Bei einem Leistungs- oder Intelligenztest wird z.B. erfasst, welche bzw. wie viele Aufgaben die Person richtig beantwortet hat. Bei einem Test zur Messung einer Persönlichkeitseigenschaft wird hingegen die Zustimmung und Ablehnung zu vorgegebenen Aussagen protokolliert. Als Ergebnis des Tests erhält die Person eine Schätzung ihrer Fähigkeit bzw. ihrer Ausprägung der latenten Eigenschaft.

Vor der Verwendung des fertigen Tests liegt aber die Phase der Testkonstruktion: Ein Expertenteam generiert dafür zunächst einen Satz von Aufgaben, die z.B. mathematische Kompetenzen von Grundschülern der 1. und 2. Klasse abdecken. Diese Aufgaben werden dann einer kritischen Prüfung unterzogen, und oft müssen Aufgaben aussortiert werden, weil sie bestimmte Kriterien nicht erfüllen.

Eine der häufigsten Ursachen für den Ausschluss von Aufgaben ist, dass sie nicht nur die interessierende Eigenschaft messen, sondern noch weitere Eigenschaften sich auf die Beantwortung der Frage auswirken. Bei der Entwicklung eines Tests zur Messung der mathematischen Kompetenz von Grundschülern können z.B. Aufgaben vorkommen, die – neben der mathematischen Kompetenz – auch durch die verbale Kompetenz der Schüler beeinflusst werden: Da es sich bei vielen Mathematik-Aufgaben um Textaufgaben handelt, kann die Formulierung der Aufgabe z.B. Schüler mit Deutsch als Fremdsprache benachteiligen, so dass sie in dem Test schlechter abschneiden – obwohl sie genauso gut in Mathematik sind wie ihre Mitschüler.

So können einzelne Aufgaben, die bestimmte Gruppen von Personen bevorzugen oder benachteiligen, dazu führen, dass das gesamte Testergebnis verzerrt ist und faire Vergleiche zwischen den Personen nicht mehr möglich sind.

Um das zu verhindern, müssen alle Test-Aufgaben in der Konstruktions-Phase des Tests einer strengen Prüfung unterzogen werden, z.B. mithilfe des Rasch-Modells. So können nachweislich unfaire Aufgaben aus dem Test ausgeschlossen bzw. durch andere Aufgaben ersetzt werden. Erst wenn der Test aus den restlichen Aufgaben alle Prüfungen erfüllt, ist er „fertig“ und kann sinnvoll zur Testung in der Praxis eingesetzt werden.

Aus diesem Grund – und nicht etwa, um die Leser unnötig zu quälen – nimmt die statistische Überprüfung von psychologischen Tests in den Lehrbüchern zur Testtheorie und Testkonstruktion oft mehr Raum ein als die praktische Anwendung der fertigen Tests: Nur psychologische Tests, die eine strenge Überprüfung bestanden haben, erlauben objektive Messungen und faire Vergleiche. Ohne diese Überprüfung könnte es hingegen passieren, dass ein psychologischer Test zu falschen Schlüssen führt – die wiederum zu schwerwiegenden Fehlentscheidungen über das schulische und berufliche Fortkommen einer Person oder den Behandlungsbedarf einer psychischen Erkrankung führen können. Auch wenn Sie psychologische Tests „nur“ in der Praxis anwenden wollen, ist es also wichtig, sich mit den Kriterien auseinanderzusetzen, anhand derer Sie objektive und faire Tests erkennen können.

Die Überprüfbarkeit des Rasch-Modells – und gerade auch die Tatsache, dass seine Gültigkeit für viele psychologische Tests abgelehnt werden muss – stellt dabei seine größte Stärke dar, und unterscheidet den Ansatz der Item-Response-Theorie (IRT), der neben dem Rasch-Modell noch viele Erweiterungen umfasst (von denen einige am Ende dieses Buches kurz dargestellt sind), von dem der klassischen psychologischen Testtheorie (vgl. z.B. Irtel, 1996; Steyer & Eid, 2001; Bühner, 2006; Moosbrugger & Kelava, 2007).

Dieser Gedanke – dass die Möglichkeit, ein Modell abzulehnen, eine Stärke und keine Schwäche des Modells darstellt – ist Ihnen vielleicht schon aus einer einführenden Vorlesung zur Wissenschaftstheorie vertraut: Eine wissenschaftliche Theorie ist umso stärker, je leichter sie falsifizierbar ist. Das kann man sich an einem einfachen Beispiel klarmachen: Die Theorie „Immer wenn der Hund eine Katze sieht, läuft er weg oder auch nicht.“ ist in 100% der Fälle zutreffend – und liefert gerade dadurch keinerlei Erkenntnisgewinn. Die Theorie „Immer wenn der Hund eine Katze sieht, läuft er weg.“ hingegen kann (in diesem Fall von dem betroffenen Hund) sehr einfach widerlegt werden. Stellt sie sich aber in vielen Fällen als zutreffend heraus, spricht das für ihren empirischen Gehalt, und wir sind um eine Erkenntnis über den Hund reicher.

Genauso kann es auch passieren, dass das Rasch-Modell anhand der Daten aus einem psychologischen Test widerlegt wird, d.h. dass der Test den Anforderungen des Rasch-Modells nicht genügt. Befindet man sich noch in der Phase der Testkonstruktion, lässt sich das oft noch ändern, indem man z.B. Aufgaben, die einzelne Gruppen von Personen bevorzugen oder benachteiligen, aus dem Test streicht oder durch andere Aufgaben ersetzt. Andernfalls

sollte man kritisch hinterfragen, ob man einen Test verwenden möchte, der potentiell zu Fehlschlüssen führen kann.
Diese oft als „zu streng“ empfundene Konsequenz des Rasch-Modells hat vermutlich – neben der Abschreckungswirkung seiner mathematischen Darstellung – dazu geführt, dass es von vielen Psychologen seit seiner Veröffentlichung durch den dänischen Statistiker Georg Rasch im Jahr 1960 über mehrere Jahrzehnte eher als theoretische Spielerei und Schreck von Prüflingen als als Hilfsmittel zur praktischen Testkonstruktion betrachtet wurde.
Inzwischen haben sich aber gleich zwei erfreuliche Strömungen durchgesetzt: Das Rasch-Modell wird seit einigen Jahren nicht nur erfolgreich zur Konstruktion neuer Tests verwendet – berühmtestes Beispiel ist sein Einsatz in der PISA Studie der OECD (siehe z.B. Frey, Carstensen, Walter, Rönnebeck, & Gomolka, 2006) – sondern auch, um bereits etablierte psychologische Tests nachträglich daraufhin zu überprüfen, ob sie mit dem Rasch-Modell vereinbar sind (Rost, Carstensen, & Von Davier, 1999; Bühner, Krumm, Ziegler, & Schmidt-Atzert, 2006; Shea, Tennant, & Pallant, 2009).
Entsprechend kommt inzwischen sowohl im Psychologie-Studium als auch in der diagnostischen Praxis niemand mehr um das Rasch-Modell herum – und das ist auch gut so, denn einen psychologischen Test ohne die Überprüfung seiner Objektivität zu entwickeln oder anzuwenden bedeutet nicht nur, dass man sich komplizierte Formeln erspart, sondern auch, dass man das Risiko eingeht zu falschen Schlussfolgerungen und gravierenden Fehlentscheidungen zu kommen. Insbesondere auf die Thematik von Aufgaben, die einzelne Gruppen von Personen bevorzugen oder benachteiligen, werden wir deshalb im Folgenden immer wieder zurückkommen.

# 2

# Mathematische Formulierung und inhaltliche Bedeutung des Rasch-Modells

In diesem Kapitel werden wir die Struktur der Daten, die sich z.B. aus einem Leistungstest ergeben, und das Rasch-Modell selbst kennenlernen (mit „das Rasch-Modell“ ist übrigens meist die Modellgleichung gemeint, also eine Formel). Dazu kommen statistische Annahmen und Eigenschaften des Modells, deren mathematische und inhaltliche Bedeutung erklärt wird.

Dieses Kapitel ist als Hilfestellung zur vertieften Prüfungsvorbereitung gedacht, und enthält evtl. mehr Formeln und Zwischenschritte, als Sie zum grundsätzlichen Verständnis des Verfahrens und für die praktische Anwendung benötigen (ggf. können Sie also die Formeln beim Lesen überspringen; die wichtigsten Formeln sind durch Umrandungen hervorgehoben).

## 2.1 Die Datenmatrix

Mit dem Rasch-Modell können Tests zur Leistungs- und Einstellungsmessung ausgewertet werden. Ein Test soll eine latente (d.h. nicht direkt beobachtbare) Fähigkeit oder Einstellung der Personen messen. Für jede Person, die an dem Test teilnimmt, wird dabei notiert, ob die Person eine Aufgabe richtig oder falsch beantwortet hat, bzw. ob die Person einer Aussage zustimmt oder sie ablehnt.
Diese Antworten lassen sich dann in einer einfachen Tabelle oder Matrix zusammenfassen: Für jede richtige Antwort oder Zustimmung erhält die Person eine 1, für jede falsche Antwort oder Ablehnung eine 0. Die in Tabelle 2.1 dargestellte Datenmatrix enthält z.B. die Antworten von 4 Personen auf 6 Aufgaben bzw. Fragen. Die erste Person hat z.B. die zweite, vierte und sechste Aufgabe richtig beantwortet. (Der Einfachheit halber werden wir im Folgenden als Beispiel immer von einem Leistungstest ausgehen, der die Fähigkeit einer Person misst.)
Allgemein betrachten wir die Situation, dass $i = 1, \ldots, n$ Personen an einem Test mit $j = 1, \ldots, m$ Aufgaben teilnehmen. In unserem Beispiel wäre $n = 4$

**Tabelle 2.1.** Datenmatrix eines Tests mit 6 Aufgaben, an dem 4 Personen teilgenommen haben.

| | Aufgabe | | | | | |
|---|---|---|---|---|---|---|
| Person | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 2 | 0 | 1 | 1 | 0 | 1 | 1 |
| 3 | 0 | 1 | 1 | 1 | 0 | 0 |
| 4 | 1 | 0 | 0 | 1 | 0 | 0 |

und $m = 6$. Die Datenmatrix besteht also aus $n$ Zeilen und $m$ Spalten. Die Einträge der Datenmatrix bezeichnen wir mit $u_{ij}$, d.h. der Eintrag $u_{ij}$ gehört zur Person $i$ (liegt also in der $i$-ten Zeile) und zur Aufgabe $j$ (also in der $j$-ten Spalte – dabei wird immer zuerst die Zeile, dann die Spalte angegeben). Die Antwort der ersten Person auf die zweite Aufgabe heißt also $u_{12}$. Die Datenmatrix in dieser allgemeinen Schreibweise ist in Tabelle 2.2 dargestellt.

**Tabelle 2.2.** Datenmatrix eines Tests mit $m$ Aufgaben, an dem $n$ Personen teilgenommen haben.

| | Aufgabe | | | | | | |
|---|---|---|---|---|---|---|---|
| Person | 1 | 2 | ... | $j$ | ... | $m-1$ | $m$ |
| 1 | $u_{1,1}$ | $u_{1,2}$ | ... | $u_{1,j}$ | ... | $u_{1,m-1}$ | $u_{1,m}$ |
| 2 | $u_{2,1}$ | $u_{2,2}$ | ... | $u_{2,j}$ | ... | $u_{2,m-1}$ | $u_{2,m}$ |
| $\vdots$ | ... | ... | ... | ... | ... | ... | ... |
| $i$ | $u_{i,1}$ | $u_{i,2}$ | ... | $u_{i,j}$ | ... | $u_{i,m-1}$ | $u_{i,m}$ |
| $\vdots$ | ... | ... | ... | ... | ... | ... | ... |
| $n-1$ | $u_{n-1,1}$ | $u_{n-1,2}$ | ... | $u_{n-1,j}$ | ... | $u_{n-1,m-1}$ | $u_{n-1,m}$ |
| $n$ | $u_{n,1}$ | $u_{n,2}$ | ... | $u_{n,j}$ | ... | $u_{n,m-1}$ | $u_{n,m}$ |

Eine zentrale Idee der Item-Response-Theorie ist, dass das Ergebnis einer Person in einem Test nicht deterministisch von ihrer Fähigkeit abhängt, sondern dass dabei auch immer etwas Zufall im Spiel ist: Man kann sich gut vorstellen, dass eine Person (obwohl ihre Fähigkeit ja immer gleich bleibt) nicht genau dieselbe Punktzahl erreichen würde, wenn sie an zwei unterschiedlichen Tagen an demselben Test teilnehmen würde. An einem Tag wäre sie vielleicht müde und würde viele Flüchtigkeitsfehler machen. An einem anderen Tag wäre sie hingegen in Top-Form und würde sogar bei einigen Aufgaben richtig raten, deren Lösung sie eigentlich nicht weiß.

Entsprechend werden wir später oft von der Wahrscheinlichkeit sprechen, mit der eine Person eine Aufgabe richtig beantwortet, der sogenannten Lösungswahrscheinlichkeit. Eine sehr fähige Person hat z.B. eine hohe Wahrscheinlichkeit, eine Aufgabe richtig zu beantworten – trotzdem kann man nicht 100%-ig sicher sein, ob sie die Aufgabe wirklich lösen wird, bevor sie es versucht hat.
Deshalb gibt es neben den Einträgen $u_{ij}$ in der Datenmatrix, die wir schon beobachtet haben (d.h. wir wissen bereits, ob die Person die Aufgabe gelöst hat oder nicht, also ob an der Stelle in der Datenmatrix eine 1 oder eine 0 steht), noch die Zufallsvariablen $U_{ij}$. Die Zufallsvariablen $U_{ij}$ stehen sozusagen für das noch unbekannte Ergebnis bevor die Personen die Aufgaben bearbeiten. Im nächsten Abschnitt brauchen wir diese Unterscheidung, um die Wahrscheinlichkeit, dass ein bestimmtes Ergebnis (0 oder 1) eintritt, in einer allgemeinen Formel darzustellen, die für beide möglichen Ergebnisse (0 oder 1) gilt. Die Wahrscheinlichkeit, dass Person $i$ bei der Beantwortung von Aufgabe $j$ das Ergebnis $u_{ij}$ erzielt, nennen wir dann $P(U_{ij} = u_{ij})$ ($P$ steht dabei für das englische Wort „Probability“).

## 2.2 Die Modellgleichung

Die Wahrscheinlichkeit, dass eine Person eine Aufgabe richtig beantwortet, hängt natürlich von der Fähigkeit der Person ab: Eine fähigere Person wird die Aufgabe eher lösen als eine weniger fähige Person – trotzdem kann auch die fähigere Person mal einen Flüchtigkeitsfehler begehen, wie im letzten Abschnitt besprochen.
Zusätzlich hängt die Wahrscheinlichkeit, dass eine Person eine Aufgabe richtig beantwortet, aber auch von der Schwierigkeit der Aufgabe selbst ab: Wenn die Aufgabe sehr leicht ist, wird sie auch eine weniger fähige Person mit hoher Wahrscheinlichkeit lösen können. Das heißt, dass sowohl die Fähigkeit der Person als auch die Schwierigkeit der Aufgabe in der Modellgleichung berücksichtigt werden sollten.
Die Modellgleichung soll also die Wahrscheinlichkeit beschreiben, dass eine Person mit einer bestimmten Fähigkeit eine Aufgabe mit einer bestimmten Schwierigkeit richtig beantwortet. Aus dieser Überlegung können wir schon einige Eigenschaften ableiten, die die Modellgleichung haben soll:

1. Die Fähigkeit der Person soll berücksichtigt werden (wir benennen die Fähigkeit der Person $i$ mit $\theta_i$).
2. Die Schwierigkeit der Aufgabe soll berücksichtigt werden (wir benennen die Schwierigkeit der Aufgabe $j$ mit $\beta_j$).
3. Je fähiger die Person, desto höher ihre Lösungswahrscheinlichkeit – d.h. die Modellgleichung soll eine Funktion sein, die mit der Personen-Fähigkeit ansteigt.
4. Da es sich um eine Wahrscheinlichkeit handelt, muss sie aber zwischen den Grenzen 0 und 1 bleiben.

Das Rasch-Modell (Formel 2.1) erfüllt alle diese sinnvollen Forderungen – und hat, wie wir später sehen werden, noch einige andere angenehme Eigenschaften. Die Modellgleichung für das Rasch-Modell lautet:

$$\boxed{P(U_{ij} = 1|\theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \qquad (2.1)}$$

Wenn wir diese Formel stückchenweise betrachten, finden wir alle geforderten Eigenschaften wieder: Die Lösungswahrscheinlichkeit hängt

1. von der Personen-Fähigkeit $\theta_i$ und
2. der Aufgaben-Schwierigkeit $\beta_j$ ab.

Das erkennt man daran, dass die beiden Parameter $\theta_i$ und $\beta_j$ auf der linken Seite der Gleichung als Bedingung hinter dem senkrechten Strich stehen (d.h. die Lösungswahrscheinlichkeit ist eine bedingte Wahrscheinlichkeit, die von der Personen-Fähigkeit und der Aufgaben-Schwierigkeit abhängt) und auf der rechten Seite der Gleichung in der Form $\theta_i - \beta_j$ auftauchen.
An dieser Differenz sieht man, dass die Lösungswahrscheinlichkeit wie gewünscht davon abhängt, wie groß die Fähigkeit der Person im Vergleich zur Schwierigkeit der Aufgabe ist: Ist die Person fähiger als die Aufgabe schwer ist (ist also $\theta_i > \beta_j$), ergibt sich bei diesem Vergleich eine positive Differenz; ist aber die Aufgabe schwerer als die Person fähig ist (ist also $\theta_i < \beta_j$), ergibt sich eine negative Differenz.
Diese Differenz kommt gleich an zwei Stellen in der Formel vor: im Zähler und im Nenner des Bruches. Der kompliziert aussehende Bruch

$$\frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

beschreibt aber eine ganz einfache Funktion. Diese Funktion mit der vereinfachten Form

$$\frac{e^x}{1 + e^x}$$

nennt man logistische Funktion[①].

① Die logistische Funktion wird auch in anderen Bereichen der Statistik angewendet, z.B. bei der logistischen Regression mit einer binären Zielgröße. Als Alternative zur logistischen Funktion kann man auch andere S-förmige Kurven verwenden, z.B. die kumulierte Dichte der Normalverteilung. In der Regression unterscheidet man entsprechend das „Logit Modell" (logistische Funktion) und das „Probit Modell" (kumulierte Dichte der Normalverteilung). Da die beiden Kurven kaum zu unterscheiden sind, führen beide Ansätze meist zu sehr ähnlichen Ergebnissen. Auch im Rasch-Modell könnte man im Prinzip eine andere S-förmige Kurve wie die kumulierte Dichte der Normalverteilung verwenden – allerdings wäre diese Funktion viel komplizierter (was u.a. dazu führen würde, dass es keine einfachen suffizienten Statistiken wie im Rasch-Modell mehr gäbe, vgl. Abschnitt 2.5.1).

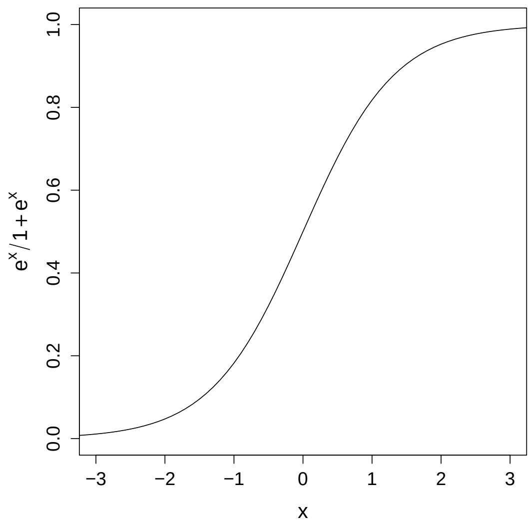


**Abb. 2.1.** Logistische Funktion.

Die logistische Funktion erinnert an ein leicht zur Seite gekipptes S, wie in Abbildung 2.1 dargestellt. Praktischerweise erfüllt diese Funktion mit ihrer S-Form gleich zwei von unseren Forderungen:

3. Die Funktion steigt, wenn der Wert $x$ im Exponenten größer wird (von links nach rechts in Abbildung 2.1).
   Da an Stelle von $x$ im Rasch-Modell die Differenz $\theta_i - \beta_j$ steht, heißt das: Die Lösungswahrscheinlichkeit steigt mit der Überlegenheit der Person über die Aufgabe, ausgedrückt in der Differenz $\theta_i - \beta_j$.
   Ist die Person z.B. fähiger als die Aufgabe schwer ist (also $\theta_i > \beta_j$ und daher $\theta_i - \beta_j > 0$), ist die Lösungswahrscheinlichkeit groß. Ist die Person hingegen weniger fähig als die Aufgabe schwer ist (also $\theta_i < \beta_j$ und daher $\theta_i - \beta_j < 0$), ist die Lösungswahrscheinlichkeit klein.
4. Die Werte, die die logistische Funktion annehmen kann, liegen zwischen 0 und 1 (vgl. Wertebereich der y-Achse in Abbildung 2.1). Deshalb ist sie gut dafür geeignet, Wahrscheinlichkeiten abzubilden – wie die Lösungswahrscheinlichkeit im Rasch-Modell.

Durch den beschränkten Wertebereich ergibt sich noch eine Besonderheit in der Form der logistischen Funktion: Sie hat im mittleren Bereich die größte Steigung und flacht an den Enden ab. Auf die Stärke der Steigung im mittleren Bereich der Funktion werden wir im nächsten Abschnitt zurückkommen, wenn es um die Trennschärfe von Aufgaben geht.

## 2.3 Aufgaben- und Personencharakteristische Kurven

Die logistische Funktion in Abbildung 2.1 wird im Rasch-Modell verwendet, um den Verlauf der Lösungswahrscheinlichkeit für eine Aufgabe in Abhängigkeit von der Personen-Fähigkeit darzustellen. Diese Darstellung bezeichnet man als aufgabencharakteristische Kurve oder ICC (vom englischen „Item Characteristic Curve“).
An einer ICC (vgl. Abbildung 2.2 für eine Aufgabe mit Schwierigkeit $\beta_j = 6$) kann man ablesen, dass z.B. eine Person $i$ mit Fähigkeit $\theta_i = 8$ (auf der x-Achse ablesen) eine sehr hohe Lösungswahrscheinlichkeit hat (auf der y-Achse ablesen), weil die Person fähiger ist als die Aufgabe schwer ist. Eine Person mit Fähigkeit $\theta_i = 4$ hat hingegen nur eine geringe Lösungswahrscheinlichkeit, weil die Person weniger fähig ist als die Aufgabe schwer ist.
Man sieht auch, dass die Lösungswahrscheinlichkeit genau 50% beträgt, wenn die Person genauso fähig ist wie die Aufgabe schwer ist, also $\theta_i = \beta_j = 6$. In diesem Fall stehen also die Chancen 50/50, dass die Person die Aufgabe „besiegt“, oder die Aufgabe die Person „besiegt“.

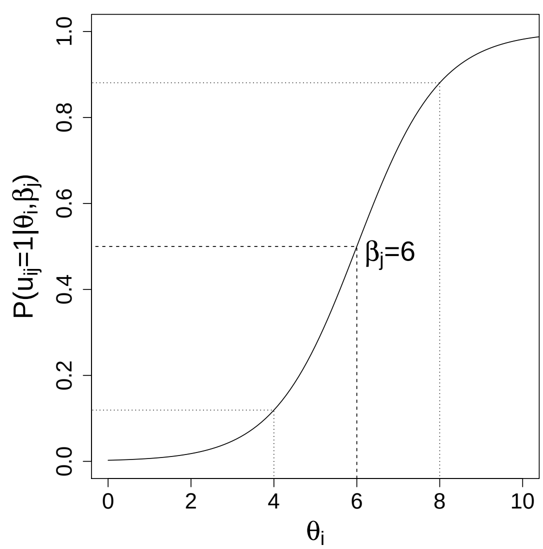


**Abb. 2.2.** Aufgabencharakteristische Kurve (ICC) für eine Aufgabe mit Schwierigkeit $\beta_j = 6$.

Ein echter Test besteht natürlich aus mehr als einer Aufgabe. Deshalb zeichnet man oft die ICCs von allen Aufgaben nebeneinander, wie in Abbildung 2.3. Bei dieser Darstellung fällt auf, dass beim Rasch-Modell alle ICCs parallel zueinander verlaufen: Die ICCs sind zwar nach links oder rechts verschoben, je nachdem ob die Aufgabe leichter oder schwieriger ist, aber die Form ist immer dieselbe.

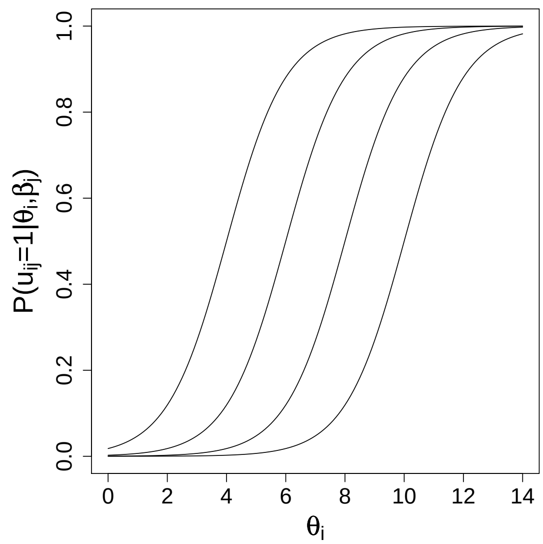


**Abb. 2.3.** Aufgabencharakteristische Kurven (ICCs) für mehrere Aufgaben.

Diese Eigenschaft ergibt sich aus der Modellgleichung des Rasch-Modells: Darin gibt es für jede Aufgabe nur einen Parameter, also einen Wert der sich ändern kann, nämlich ihre Schwierigkeit $\beta_j$. Es gibt nicht noch zusätzlich für jede Aufgabe einen Parameter, der die Steigung oder Form der Kurve verändern könnte; deshalb haben alle Aufgaben dieselbe Steigung.
Aber auch inhaltlich bedeutet das, dass alle Aufgaben etwas gemeinsam haben müssen, wenn man das Rasch-Modell verwenden will: Die Steigung im mittleren Bereich der ICCs bezeichnet man auch als Trennschärfe. Je höher die Trennschärfe einer Aufgabe, desto genauer kann man mithilfe dieser Aufgabe zwischen Personen mit unterschiedlichen Fähigkeiten unterscheiden.
Wie in Abbildung 2.4 dargestellt, ergeben sich bei einer Aufgabe mit hoher Trennschärfe (links) für zwei Personen mit leicht unterschiedlichen Fähigkeiten (auf der x-Achse) deutlich unterschiedliche Lösungswahrscheinlichkeiten (auf der y-Achse), wohingegen sich bei einer Aufgabe mit niedriger Trennschärfe (rechts) nur ein geringer Unterschied in den Lösungswahrscheinlichkeiten ergibt. Die zweite Aufgabe ist also weniger gut geeignet, um zwischen den beiden Personen zu unterscheiden.

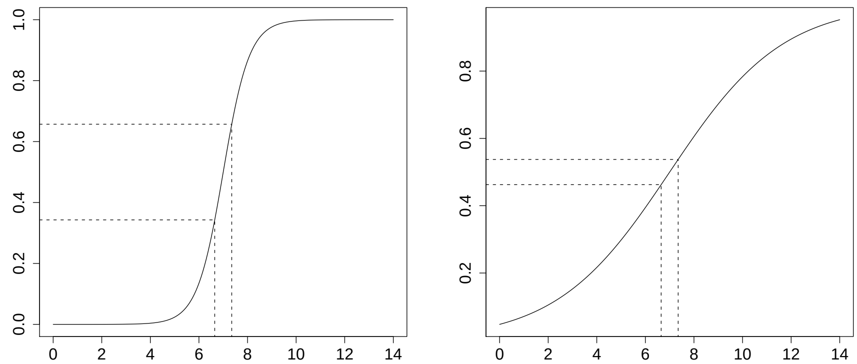


**Abb. 2.4.** ICCs für zwei Aufgaben mit hoher (links) und niedriger (rechts) Trennschärfe.

Die Tatsache, dass die Modellgleichung des Rasch-Modells nur Aufgaben mit gleicher Trennschärfe zulässt, kann man als Eigenschaft des Modells betrachten – aber auch als eine strenge Forderung, die jeder Test erfüllen muss, der mit dem Rasch-Modell ausgewertet werden soll. Bei einem echten Test kann es nämlich durchaus passieren, dass er Aufgaben enthält, die unterschiedlich trennscharf sind – dieser Test ist dann nicht „Rasch-skalierbar“, d.h. er erlaubt keine objektiven Messungen im Sinne von Rasch (1960) (vgl. Kapitel 2.5.3).
In späteren Kapiteln werden wir sowohl statistische Tests kennenlernen, mit denen sich überprüfen lässt, ob die Daten aus einem echten Test den strengen theoretischen Annahmen des Rasch-Modells genügen, als auch Modelle, die unterschiedlich trennscharfe Aufgaben zulassen.

## 2.4 Unterschiedliche Darstellungen der Modellgleichung

Die Modellgleichung für das Rasch-Modell, wie sie in Formel 2.1 dargestellt ist, lässt sich auf verschiedene Weise umformen. All diese unterschiedlichen Formeln beschreiben immer dasselbe Modell, aber Sie sollten sie einmal gesehen haben, damit Sie sie in den späteren Kapiteln oder anderen Büchern wiedererkennen.
Bisher haben wir immer die Wahrscheinlichkeit, dass eine Person eine Aufgabe löst $P(u_{ij} = 1|\theta_i, \beta_j)$ betrachtet (Formel 2.1). Natürlich gibt es aber auch den umgekehrten Fall, dass die Person die Aufgabe nicht löst.
Da es nur diese beiden Möglichkeiten gibt – die Person löst die Aufgabe, oder sie löst sie nicht – sind die beiden zugehörigen Wahrscheinlichkeiten sogenannte Gegenwahrscheinlichkeiten, die sich zu eins aufaddieren:

$$P(u_{ij} = 1|\theta_i, \beta_j) + P(u_{ij} = 0|\theta_i, \beta_j) = 1$$

Die Gegenwahrscheinlichkeit $P(u_{ij} = 0|\theta_i, \beta_j)$, dass die Person die Aufgabe nicht löst, ergibt sich also aus unserer ursprünglichen Formel für $P(u_{ij} = 1|\theta_i, \beta_j)$ durch:

$$\begin{aligned} P(u_{ij} = 0|\theta_i, \beta_j) &= 1 - P(u_{ij} = 1|\theta_i, \beta_j) \\ &= 1 - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \\ &\overset{①}{=} \frac{1 + e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \\ &= \frac{1 + e^{\theta_i - \beta_j} - e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \\ &= \frac{1}{1 + e^{\theta_i - \beta_j}} \end{aligned}$$

Nun kennen wir für jedes der beiden möglichen Ergebnisse – die Person löst die Aufgabe oder sie löst sie nicht – die Formel für die entsprechende Wahrscheinlichkeit. Praktischerweise lassen sich diese beiden Formeln aber auch in einer gemeinsamen Formel zusammenfassen.

$$\boxed{P(U_{ij} = u_{ij}|\theta_i, \beta_j) = \frac{e^{u_{ij} \cdot (\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \qquad (2.2)}$$

In dieser Formel ist noch offen gelassen, ob sich als Ergebnis $u_{ij}$ der Wert 1 oder 0 ergibt. (Diese Formel werden wir später zur Schätzung der Personen- und Aufgaben-Parameter verwenden.) Aber auch die zwei einzelnen Formeln für $P(u_{ij} = 1|\theta_i, \beta_j)$ und $P(u_{ij} = 0|\theta_i, \beta_j)$ ergeben sich direkt aus Formel 2.2, wenn man für $u_{ij}$ jeweils 1 oder 0 einsetzt:

$$\frac{e^{u_{ij} \cdot (\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} = \begin{cases} \dfrac{e^{1 \cdot (\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} = \dfrac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} & \text{für } u_{ij} = 1 \\[2ex] \dfrac{e^{0 \cdot (\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \overset{②}{=} \dfrac{1}{1 + e^{\theta_i - \beta_j}} & \text{für } u_{ij} = 0 \end{cases}$$

① Um in der nächsten Zeile beide Terme zu einem Bruch mit einem gemeinsamen Nenner zusammenfassen zu können, wird für die 1 ein künstlich erweiterter Bruch eingesetzt: $\frac{1+e^{\theta_i - \beta_j}}{1+e^{\theta_i - \beta_j}} = 1$

② $e^0 = 1$