PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 427

Taksonomia 27
Klasyfikacja i analiza danych – teoria i zastosowania

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania", zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 427 • 2016

Taksonomia 27 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Justyna Brzezińska
Uniwersytet Ekonomiczny w Katowicach
e-mail: justyna.brzezinska@ue.katowice.pl

MODELE IRT I MODELE RASCHA W BADANIACH TESTOWYCH

IRT AND RASCH MODELS IN TEST MEASUREMENT

DOI: 10.15611/pn.2016.427.05

Streszczenie: Modele IRT należą do coraz bardziej rozwijających się modeli w badaniach psychologicznych, medycznych, marketingowych, czy społecznych. Jednym z modeli należącym do rodziny IRT (*Item Response Theory*), czyli modeli teorii odpowiedzi na pozycje testowe, jest model Rascha, w którym istnieją dwa parametry: jeden związany z cechami przedmiotu badania, drugi natomiast z cechami sytuacji, w której przeprowadzane jest badanie. W celu estymacji parametrów pozycji wykorzystuje się metodę największej wiarygodności. Modele te wykorzystywane są zazwyczaj w badaniach edukacyjnych czy psychologicznych. Celem artykułu jest prezentacja różnych rodzajów modeli IRT w zależności od liczby parametrów, a także szczegółowy opis modelu Rascha. Dodatkowo w artykule zaprezentowana zostanie analiza IRT dla danych binarnych z wykorzystaniem pakietu `ltm` oraz funkcji: `rasch, ltm, tpm` programu **R**.

Słowa kluczowe: modele IRT, modele Rascha, analiza teorii odpowiedzi na pozycje, modele ze zmiennymi ukrytymi.

Summary: Item Response Theory (IRT) is an extension of Classical Test Theory (CCT) and focuses on how specific test items function in assessing construct. They are widely known models in psychological, medical, marketing and social sciences. One of the most popular IRT models is Rasch model used to separate the ability of test takers and the quality of the test. The main characteristic of IRT models, the Rasch model being the most prominent, concerns the separation of two kinds of parameters: one that describes qualities of subjects under investigation, the other relates to qualities of the situation under which the response of a subject is observed. Maximum likelihood estimation is used for parameter estimation. In this paper we present IRT analysis for binary data with the use of `ltm` package with `rasch, ltm, tpm` functions in **R**.

Keywords: IRT models, Rasch model, measurement theory, latent variables models.

1. Wstęp

Analiza pozycji testowych jest istotnym elementem procesu konstruowania testów. Ograniczenia klasycznych indeksów jakości pozycji testowych od dawna utrudniają analizę i interpretację danych uzyskanych z kwestionariuszy testowych. Zastosowanie modeli IRT w analizie pozycji testowych pozwala przezwyciężyć ograniczenia klasycznych wskaźników jakości zadań testowych.

W psychometrycznej teorii pomiaru znane są dwa podejścia: klasyczna teoria testu (CCT, *Classical Test Theory*) oraz współczesne modele teorii odpowiedzi na pozycje (IRT, *Item Response Theory*). Klasyczna teoria testu, która powstała na początku XX w., była przez długi czas pod wpływem korelacyjnej teorii Spearmana, co jest odzwierciedlone w kluczowej roli, jaką odgrywała koncepcja rzetelności testu. W statystyce i w rachunku prawdopodobieństwa dominowało wówczas podejście częstościowe. Dopiero w 1968 r. F.M. Lord i M.R. Novick dokonali przeformułowania klasycznej teorii testu na taką, która jest w zgodzie z aksjomatyzacją teorii prawdopodobieństwa zaproponowaną przez Kołmogorowa [Lord, Novick 1968].

Teoria odpowiedzi na pozycje opierała się natomiast na osiągnięciach Fischera. Początkowo model IRT określał relację pomiędzy ciągłą zmienną ukrytą a dychotomicznymi pozycjami testowymi [Birnbaum 1968; Lord, Novick 1968; Rasch 1960, 1966, 1977]. W latach późniejszych został on rozszerzony poprzez uwzględnienie pozycji politomicznych [Bock 1972; Samejima 1969]. Teoria odpowiedzi na pozycje (IRT) jest psychometrycznym terminem używanym do określenia pewnej rodziny modeli, opisujących sposób udzielania przez badane osoby odpowiedzi na poszczególne pozycje testów (*items*). W probabilistycznej teorii testu bada się prawdopodobieństwo reakcji na stwierdzenie, będące funkcją zmiennej ukrytej określającej poziom umiejętności (*ability*), na której dokonany jest pomiar cechy i pozycji parametrów będących poziomem trudności (*difficulty*). Teoria IRT dostarcza narzędzi statystycznych pozwalających analizować zachowania ucznia w stosunku do pojedynczego zadania testowego, a nie całego testu.

Ideą teorii odpowiedzi na pozycje testowe jest stworzenie modelu statystycznego określającego rozkład odpowiedzi na pozycje testu w ramach pewnej zmiennej ukrytej, reprezentującej poziom mierzonej testem cechy. Cel ten jest osiągany poprzez wprowadzenie założenia o jednowymiarowości testu oraz zdefiniowanie rodziny dopuszczalnych krzywych charakterystycznych (ICC, *Item Characteristic Curve*) pozycji testowych opisujących zależność rozkładu poszczególnych pozycji przy ustalonym poziomie zmiennej ukrytej. Krzywe są obrazem związku pomiędzy prawdopodobieństwem udzielenia przez jednostkę poprawnej odpowiedzi (czy też odpowiedzi zgodnej z kluczem) a różnymi wartościami cechy ukrytej. Krzywe te przybierają kształt litery *S* i tworzone są przez modelowanie prawdopodobieństwa sukcesu z wykorzystaniem modelu logistycznego. Wyższym wartościom zdolności respondenta odpowiadają wyższe wartości prawdopodobieństwa sukcesu określającego prawdopodobieństwo poprawnej odpowiedzi.

2. Charakterystyka modeli IRT oraz modelu Rascha

Analiza pozycji obejmuje: klasyczną teorię testów (*Classical Test Theory*) oraz modele cech ukrytych (*Latent Trait Models*). Modele cech ukrytych dzielą się z kolei na modele teorii odpowiedzi na pozycje (*Item Response Theory models*) oraz na modele Rascha, które są równoważne jednoparametrycznemu modelowi IRT.

Model IRT opisuje rozkład prawdopodobieństwa wektora odpowiedzi na zadania $\mathbf{U} = (U_1, U_2, \ldots, U_n)$ dla jednostki obserwacji, wylosowanej z populacji k:

$$P(\mathbf{U} = \mathbf{u} | k) = \int f(\mathbf{u}, \theta, \boldsymbol{\beta}) \psi_k(\theta) d\theta, \tag{1}$$

gdzie: θ – losowa zmienna ukryta opisująca poziom mierzonej umiejętności; $\psi_k(\theta)$ – funkcja gęstości prawdopodobieństwa określająca rozkład zmiennej θ w populacji k; $f(\mathbf{u}, \theta, \boldsymbol{\beta})$ – funkcja określająca prawdopodobieństwo zaobserwowania konkretnej wartości $\mathbf{u}$ wektora odpowiedzi $\mathbf{U}$, w zależności od poziomu umiejętności θ oraz wektora parametrów $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$, gdzie parametry zadania $\boldsymbol{\beta}_i$ także mogą przybierać postać wektora.

Podstawowym założeniem modeli IRT jest sfaktoryzowanie funkcji określającej prawdopodobieństwo całego wektora odpowiedzi $f(\mathbf{u}, \theta, \boldsymbol{\beta})$ do iloczynu tzw. funkcji charakterystycznych poszczególnych zadań:

$$f(\mathbf{u}, \theta, \boldsymbol{\beta}) = \prod_{i=1}^{n} f_i(u_i, \theta, \boldsymbol{\beta}_i). \tag{2}$$

Założenie (2) nosi nazwę lokalnej niezależności, która mówi, że gdy poziom umiejętności θ jest znany, wówczas odpowiedzi na zadania testu są względem siebie statystycznie niezależne i poziom umiejętności θ wystarcza do wyjaśnienia wszystkich obserwowalnych współzależności między zadaniami. Model IRT pozwala prognozować prawdopodobieństwo tego, że dana osoba odpowie w określony sposób na zadaną pozycję w teście.

Najprostszym modelem IRT jest jednoparametryczny model logistyczny tzw. model Rascha (1PL), który określa prawdopodobieństwo j-tej odpowiedzi na i-tą pozycję testową [Rasch 1960]:

$$P_{ij}\left(\theta_j, b_i\right) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}, \tag{3}$$

gdzie b_i ($-\infty < b_i < \infty$) jest parametrem trudności pozycji tzw. parametrem lokalizacji (*location, threshold parameter*), który wskazuje punkt na skali umiejętności ucznia θ, w którym prawdopodobieństwo udzielenia poprawnej odpowiedzi na zadanie wynosi 0,5. Im trudniejsze zadania, tym większy wymagany poziom umiejętności ucznia, by szansa udzielenia poprawnej odpowiedzi na zadanie wyniosła 0,5.

Model Rascha jest szczególnym przykładem modelu dwuparametrycznego (4), w którym wartość dyskryminacji wszystkich zadań jest równa jedności [Kondratek, Pokropek 2013].

Model Rascha jest z powodzeniem stosowany w praktyce badawczej od ponad 50 lat, m.in. w badaniach edukacyjnych [Scheerens 2003], medycznych [Christensen, Kreiner, Mesbah 2013] czy marketingowych [Bechtel 1985].W literaturze polskiej modele te są ciągle niedocenione, a opracowania na ich temat niewyczerpujące. Można znaleźć kilka wzmianek dotyczących modelu Rascha w badaniach ekonomicznych [Brzezińska 2015] czy też socjologicznych [Węziak 2006]. Model Rascha traktowany jest często jako szczególny przypadek bardziej ogólnych modeli IRT [Birnbaum 1968]. Choć z matematycznego punktu widzenia jednoparametryczny model IRT (*one-parameter logistic model*) jest równoważny modelowi Rascha, to zwolennicy tego ostatniego wskazują na liczne różnice w podejściach teoretycznych, stojących za tymi modelami [Andrich 2004; Masters 1982; Wright 1992, 1997].

Drugim modelem z grupy modeli IRT jest model dwuparametryczny Birnbauma (2PL), określający prawdopodobieństwo udzielenia poprawnej odpowiedzi, w zależności od poziomu umiejętności. Model ten jest określony następującym równaniem:

$$P_{ij}\left(\theta_j, a_i, b_i\right) = \frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)}, \quad (4)$$

gdzie dodatkowym, drugim parametrem modelu jest parametr dyskryminacji pozycji a_i ($-\infty < a_i < \infty$) (*slope*, *discrimination parameter*). Dyskryminacja pokazana jest jako poziom nachylenia krzywej; im krzywa jest bardziej stroma, tym silniejsza dyskryminacja pozycji, a tym samym tym większa zdolność zadania do rozróżniania poziomu umiejętności uczniów znajdujących się po obu stronach od danego punktu. Formalna ocenę dyskryminacji, w zależności od wartości parametru dyskryminacji zaproponował Baker [1985].

Kolejnym modelem z grupy modeli IRT jest model trójparametryczny Birnbauma (3PL), określony następującym równaniem:

$$P_{ij}\left(\theta_j, a_i, b_i, c_i\right) = c_i + (1 - c_i)\frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)}, \quad (5)$$

w którym parametr c_i ($0 \le c_i \le 1$) to pseudoparametr zgadywania (*guessing parameter*) pełniący funkcję dolnej asymptoty pozycji testowej. Model ten okazuje się szczególnie przydatny do modelowania odpowiedzi na zadania wyboru, w których istnieje możliwość odgadnięcia poprawnej odpowiedzi. Jest on uogólnieniem modelu dwuparametrycznego (4) w taki sposób, aby dolna asymptota przypadała powyżej zera.

Znany jest także logistyczny model czteroparametryczny (4PL) zdefiniowany równaniem:

$$P_{ij}\left(\theta_j, a_i, b_i, c_i, d_i\right) = c_i + (d_i - c_i)\frac{\exp a_i(\theta_j - b_i)}{1 + \exp a_i(\theta_j - b_i)}, \tag{6}$$

w którym parametr d_i ($0 \leq d_i \leq 1$) to parametr nieostrożności (*carelessness*), pełniący funkcję górnej asymptoty pozycji testowej. Modele te są jednak rzadko wykorzystywane w praktyce, brak ich także w znanym oprogramowaniu Mplus.

Modele IRT opierają się na kilku założeniach. Pierwszym z nich jest jednowymiarowość cechy ukrytej, która oznacza, że prawdopodobieństwo udzielenia odpowiedzi na daną pozycję testową jest funkcją pojedynczej własności charakteryzującą badaną osobę. Drugie założenie to lokalna niezależność, założenie pozwalające na oszacowanie parametrów, a trzecie to niezależność parametrów pozycji testowej. Ostatnie założenie IRT mówi o tym, iż zmienna ukryta szacowana jest na podstawie modelu. W celu estymacji parametrów modelu wykorzystuje się metodę największej wiarygodności (*MLE*). Estymatory te są zgodne, asymptotycznie normalne oraz asymptotycznie efektywne.

Istotnym czynnikiem, jaki należy rozważyć przy podejmowaniu decyzji o wyborze modelu, jest liczebność dostępnej próby, na której będzie przeprowadzana kalibracja testu. W warunkach nieograniczonych liczebnością badanej próby model trzyparametryczny będzie w zdecydowanej większości przypadków najlepszym rozwiązaniem. Model posiadający największą liczbę parametrów będzie gwarantował najlepsze dopasowanie do danych, a w związku z tym, najwyższą precyzję pomiaru. Jednak w rzeczywistości badacz rzadko dysponuje nieograniczoną możliwością wyboru liczebności grupy. Im mniejsza próba, tym oszacowania parametrów zadań są mniej dokładne, co w konsekwencji pogarsza oszacowanie poziomu umiejętności uczniów. F.M. Lord [1980] wskazywał na to, że bardziej precyzyjne pomiary dla mało licznych prób uzyskuje się za pomocą modelu jednoparametrycznego, a nie modeli bardziej złożonych, nawet gdy proces odpowiedzi na zadania wyraźnie odzwierciedla strukturę dwu- lub trzyparametryczną. Obciążone wyniki estymacji parametru dyskryminacji lub pseudozgadywania stanowią większy problem w kontekście szacowania poziomu umiejętności ucznia niż błędy spowodowane niedopasowaniem zadań do modelu IRT. Miara informacji jest w niej oparta właśnie na prawdopodobieństwie zajścia zdarzenia. Jako miarę informacji przyjmuje się wielkość niepewności usuniętą w wyniku zajścia zdarzenia (otrzymania komunikatu). Komunikaty mniej prawdopodobne dają więcej informacji.

3. Modele Rascha i ich zastosowanie w programie R

Modele Rascha w programie R dostępne są w pakiecie `ltm`, który pozwala na analizę danych zarówno binarnych, jak i porządkowych. W niniejszym artykule model Rascha z wykorzystaniem programu R zaprezentowany zostanie na przykładzie zbioru danych binarnych `LSAT` (Law School Administration Test), w którym przebadano 100 respondentów odpowiadających na 5 pytań testowych [Bock, Lieberman 1970]. Dzięki funkcji `rasch` i dodatkowemu poleceniu `constraint` oszacowano model Rascha z parametrem dyskryminacji równym 1 (*constrained model*).

Tabela 1. Poziom trudności, błąd standardowy oraz prawdopodobieństwo poprawnej odpowiedzi dla każdej z pozycji dla ograniczonego modelu Rascha z parametrem dyskryminacji równym 1

Pozycja	Poziom trudności	Błąd standardowy (SE)	Parametr dyskryminacji	$P(x = 1 \mid z = 0)$
1	2,87	0,129	1	0,946
2	1,06	0,082	1	0,743
3	0,26	0,077	1	0,564
4	1,39	0,087	1	0,800
5	2,22	0,105	1	0,902

Źródło: opracowanie własne w programie R na podstawie danych LSAT.

Ze wstępnej analizy opisowej z wykorzystaniem funkcji `descript` uzyskano następujące wartości kryteriów informacyjnych: AIC = 4956,108, BIC = 4980,646. Wartości te interpretowane są w taki sposób, że ich niższa wartość świadczy o lepszym dopasowaniu modelu do danych. Dodatkowo wiadomo, iż spośród wszystkich pozycji najtrudniejszą jest pozycja 3 (poziom trudności 0,258), a najłatwiejszą 1 (poziom trudności 2,872). Ponadto, uzyskano współczynniki korelacji chi-kwadrat w tablicy kontyngencji 2 × 2 dla wszystkich par pozycji testowych, częstości odpowiedzi przypadające na każdą pozycję testową oraz wartości alpha Cronbacha dla poszczególnych pozycji. Dzięki funkcji `coef` uzyskano wartości estymatorów parametrów, a także wartości prawdopodobieństw pozytywnej odpowiedzi na daną pozycję testową dla przeciętnego respondenta. Otrzymany porządek wskazuje uporządkowane odpowiedzi względem trudności.

Następnie oszacowano nieograniczony model Rascha (*unconstrained model*), dla którego wyniki zaprezentowano w tab. 2.

Tabela 2. Wartości estymatorów, błąd standardowy, poziom trudności oraz prawdopodobieństwo poprawnej odpowiedzi dla każdej z pozycji dla nieograniczonego modelu Rascha

Pozycja	Współczynnik	Błąd standardowy (SE)
1	–3,615	0,327
2	–1,322	0,142
3	–0,318	0,098
4	–1,739	0,169
5	–2,780	0,251
Parametr dyskryminacji	0,755	0,069

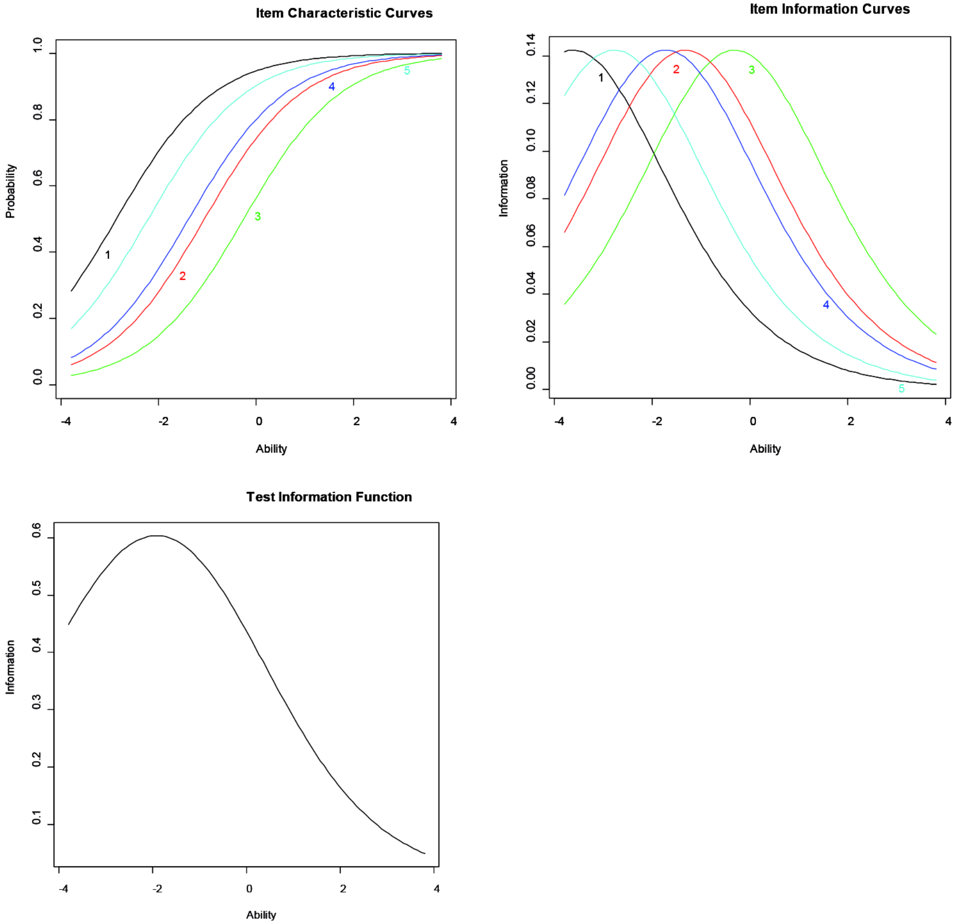
Źródło: opracowanie własne w programie R na podstawie danych LSAT.

Wartości parametru dyskryminacji są różne od 1, a wartości kryteriów informacyjnych wynoszą odpowiednio: AIC = 4945,875, BIC = 4975,322. Porównania obu modeli można dokonać dzięki funkcji `anova`, w wyniku której uzyskuje się następujące wyniki (tab. 3).

Tabela 3. Tablica ilorazu wiarygodności (*likelihood ratio table*)

Pozycja	AIC	BIC	Funkcja wiarygodności	LRT	*p-value*
Ograniczony model Rascha	4956,11	4980,65	–2473,05		
Nieograniczony model Rascha	4945,88	4975,32	–2466,94	12,23	< 0,001

Źródło: opracowanie własne w programie R na podstawie danych LSAT.



Rys. 1. Krzywa charakterystyczna, krzywa informacyjna oraz funkcja informacyjna testu dla zbioru danych LSAT

Źródło: opracowanie własne.

Wartość statystyki LRT wskazuje na to, iż model bez ograniczeń jest stosunkowo lepszy od modelu ograniczonego. Dla wybranego modelu Rascha (model nieograniczony) zaprezentowano podstawowe w analizie teorii odpowiedzi na pozycję wykresy: krzywe charakterystyczne (*Item Characteristic Curves*) oraz krzywe informacyjne (*Item Information Curves*) dla każdej z badanych pozycji oraz funkcję informacyjną (*Test Information Function*) (rys. 1).

Krzywe charakterystyczne (ICC) przedstawiają zależność między zdolnością respondentów a prawdopodobieństwem udzielenia poprawnej odpowiedzi na daną pozycję z asymptotą dolną 0, a górną 1. Każdemu zadaniu odpowiada natomiast funkcja informacyjna, która prezentuje zależność wartości informacyjnej zadań od poziomu wiedzy. Z przedstawionego wykresu funkcji informacyjnej testu widać, iż pozycje w badaniu zapewniają informację dla respondentów z niską zdolnością. Informacja dla poziomu zdolności w przedziale (–4,0) wynosi prawie 60%, a pozycja wyróżniająca respondentów z wyższym poziomem zdolności to jedynie 1/3.

4. Podsumowanie

Modele Rascha są z powodzeniem stosowane w praktyce badawczej od ponad 50 lat, m.in. w badaniach edukacyjnych [Scheerens 2003], medycznych [Christensen, Kreiner, Mesbah 2013] czy marketingowych [Bechtel 1985]. Model Rascha należy do grupy modeli, który może odwoływać się do każdego modelu pomiarowego. Modele te są różnymi sposobami parametryzacji założeń dotyczących pomiaru, sformułowanych przez G. Rascha, które powstały w obliczu potrzeby analizy danych o różnej strukturze odpowiedzi. Model Rascha jest więc kryterium, do którego można porównać dane, aby sprawdzić czy pozwalają one na obiektywny pomiar. Jest to pewnego rodzaju relacja odwrócona, w jakiej pozostają do siebie dane i model w porównaniu z podejściem IRT.

W niniejszym artykule podjęto próbę uporządkowania terminologii z zakresu modeli IRT, zaprezentowano szczegółowo model Rascha, a także podjęto próbę przedstawienia jego aplikacji w programie R w pakiecie `ltm`. Dokonano porównania modelu Rascha z ograniczeniem, w którym zakłada się, iż parametr dyskryminacji wynosi 1, z modelem Rascha bez ograniczeń. Dla modeli tych przedstawiono wartości parametrów dla każdej pozycji, ich błędy standardowe oraz kryteria informacyjne, służące ocenie dopasowania modelu do danych. Ponadto, dla wybranego modelu Rascha zaprezentowano krzywe charakterystyczne oraz krzywe informacyjne dla badanych pozycji, a także wykres funkcji informacyjnej.

Literatura

Andrich D., 2004, *Controversy and the Rasch model: A characteristic of incompatible paradigms?*, Medical Care, vol. 42, no. 1 (Supplement I–7), s. 1–16.

Baker F.B., 1985, *The basic of item response theory*, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Bechtel G.G., 1985, *Generalizing the Rasch model for consumer rating scales*, Marketing Science, vol. 4, no. 1, s. 62–73.

Birnbaum A., 1968, *Some latent trait models and their use in inferring an examinee's ability*, [w:] F.M. Lord, M.R. Novick (red.), *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, s. 395–479.

Bock R.D., 1972, *Estimating item parameters and latent ability when response are scored in two or more nominal categories*, Psychometrika, vol. 37, s. 29–51.

Bock R., Lieberman M., 1970, *Fitting a response model for n dichotomously scored items*, *Psychometrika*, vol. 35, s. 179–197.

Brzezińska J., 2015, *Rasch models in eRm package in R*, [w:] M. Papież, S. Śmiech (red.), *The 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, http://pliki.konferencjazakopianska.pl/proceedings_2015/proceedings.html, s. 29–38.

Christensen K.B., Kreiner S., Mesbah M., 2013, *Rasch Models in Health*, ISTE–Wiley, London–Hoboken.

Kondratek B., Pokropek A., 2013, *IRT i pomiar edukacyjny*, Edukacja, nr 4 (124), s. 42–66.

Lord F.M., 1980, *Applications of Item Response Theory to Practical Testing Problems*, Lawrence Erlbaum, Hillsdale.

Lord F.M., Novick M.R., 1968, *Statistical Theories of Mental Test Scores* (with contributions by a. Birnbaum), Reading, MA: Addison-Wesley.

Masters G.N., 1982, *A Rasch model for partial credit scoring*, Psychometrika, vol. 47, no. 2), s. 149–174.

Rasch G., 1960, *Probabilistic Models for some Intelligence and Attainment Tests*, Danish Institute for Education Research, Copenhagen.

Rasch G., 1966, *An individualistic approach to item analysis*, [w:] P.F. Lazarsfeld, N.W. Henry (eds.) *Readings in mathematical social sciences*, Cambridge: MIT Press, 89-107.

Rasch G., 1977, *On specific objectivity: An attempt at formalising the request for generality and validity of scientific statements*, Danish Yearbook of Philosophy, vol. 14, s. 58–94.

Samejima F., 1969, *Calibration of latent ability using response pattern of graded score*, Psychometrika Monograph Supplement, no. 17.

Scheerens J., 2003, *Educational Evaluation, Assessment, and Monitoring: A Systemic Approach*, Swets & Zeitlinger, Lisse–Exton.

Węziak D., 2006, *Zastosowanie porządkowego skalowania Rascha do optymalizacji długości skali odpowiedzi*, Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu, nr 71, *Ilościowe i jakościowe metody badania rynku*, s. 137–146.

Wright B.D., 1992, *IRT in the 1990s: Which models work best? 3PL or Rasch?*, Rasch Measurement Transactions, vol. 6, no. 1, s. 196–200.

Wright B.D., 1997, *A history of social science measurement*, Educational Measurement: Issues and Practice, vol. 16, no. 4, s. 33–45.