Procedura skalowania wyników testów

Celem badania PISA jest dostarczenie wiedzy o poziomie umiejętności uczniów m.in. w trzech głównych dziedzinach takich jak rozumowanie w naukach przyrodniczych, matematyka, czytanie ze zrozumieniem. Wszystkie badane umiejętności to cechy ukryte (latentne). Zmienne tego typu są nam dostępne jedynie poprzez ich obserwowalne konsekwencje. W przypadku badania PISA szacowanie poziomu umiejętności dla grup i podgrup uczniów odbywa się na podstawie odpowiedzi na zadania testowe wylosowanych badanych. Istotnym elementem całego procesu konstruowania testu jest procedura skalowania. Od samego początku w badaniach PISA skalowanie odbywa się poprzez aplikację teorii odpowiedzi na pozycje testowe (*IRT – Item Response Theory)*, a dokładniej uogólniona wersja modelu Rascha.

Czym jest IRT?

W najszerszym rozumieniu IRT definiuje funkcję wiążącą cechę latentną z odpowiedziami na pozycje testowe (inaczej mówiąc wskaźnikami). Pozwala to nam na oszacowanie poziomu umiejętności uczniów, czyli przypisanie im wartość na skali ukrytej.

Funkcją wiążącą jest funkcja prawdopodobieństwa odpowiedzi prawidłowej (lub nieprawidłowej) na pytanie w teście przez ucznia. Jej postać interpretować możemy w następujący sposób: im trudniejsze pytanie tym mniejsza wartość prawdopodobieństwa odpowiedzi pozytywnej oraz czym wyższe kompetencje odpowiadającego tym to prawdopodobieństwo wyższe.

Model Rascha[1]

Model Rascha to model dwuparametryczny. Pierwszym parametrem jest poziom umiejętności ucznia, a drugim poziom trudności wskaźnika. Funkcja wiążąca określa prawdopodobieństwo odpowiedzi na dany wskaźnik o pewnym poziomie trudności przez ucznia na danym poziomie umiejętności w następujący sposób:

Przez δ_i oznaczać będziemy poziom umiejętności ucznia i, a przez β_j poziom trudności wskaźnika j.

$$\mathrm{P}(\mathrm{X}_{ij} = x \mid \delta_j, \beta_i) = \frac{\exp[x(\beta_i - \delta_j)]}{1 + \exp(\beta_i - \delta_j)},$$

$$\log \frac{\mathrm{P}(\mathrm{X}_{ij} = 1 \mid \delta_j, \beta_i)}{\mathrm{P}(\mathrm{X}_{ij} = 0 \mid \delta_j, \beta_i)} = \beta_i - \delta_j$$

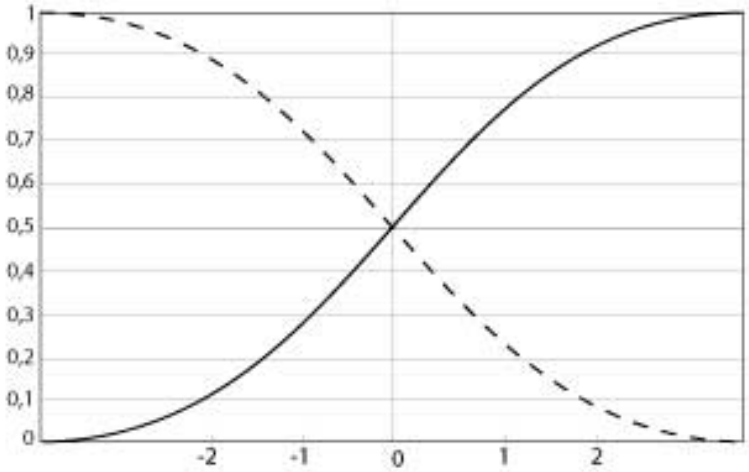
Oznacza to, że prawdopodobieństwo pozytywnej (lub negatywnej) reakcji na wskaźnik zależy od różnicy pomiędzy poziomem umiejętności ucznia i poziomem trudności pytania na jakie odpowiada. Jedyne co wpływa na wartość prawdopodobieństwa dla ucznia o danym poziomie

[1] Bardziej szczegółowa postać aplikowanego modelu znaleźć można w Raporcie Technicznym (OECD, 2005b:120).

umiejętności, to trudność pytania. W przypadku gdy odpowiedź dotyczy pytania o poziomie trudności równym jego poziomowi umiejętności prawdopodobieństwo to wynosi ½.

Graficzną interpretację powyższego zapisu przedstawiają dwie krzywe reakcji na wskaźnik. Na osi rzędnych odłożone zostało prawdopodobieństwo reakcji, a na osi odciętych różnica pomiędzy poziomem umiejętności i trudnością pytania, czyli relatywna trudność pytania. Pierwsza krzywa (ciągła) oznacza prawdopodobieństwo odpowiedzi poprawnej, a druga (przerywana) prawdopodobieństwo negatywnej.

Wykres 1. Krzywa reakcji na wskaźnik



Jedną z konsekwencji założeń modelu jest fakt, że w grupie studentów o tym samym poziomie umiejętności znajdą się wszyscy, którzy odpowiedzieli prawidłowo na tę samą liczbę pytań. Inaczej mówiąc wszyscy, którzy odpowiedzieli prawidłowo na tyle samo pytań testowych będą przypisani do tej samej klasy ukrytej. Zatem, prosta suma z liczby odpowiedzi pozytywnych będzie informować nas o tym do jakiej grupy (klasy ukrytej) należy dany uczeń, choć nie odpowiada na pytanie jaka jest jego/jej wartość na skali ukrytej.

W modelu Rascha wartości cechy ukrytej dla poziomów umiejętności uwzględniają informacje o trudności pytań. Z faktu, że wszystkie pytania możemy umieścić na skali ukrytej wynika, że całkowicie uprawnione jest porównywanie estymatorów umiejętności uczniów pomiędzy sobą, a porównanie nie zależy od poziomu trudności pytania jak byłoby w przypadku porównywania np. sumy punktów.

Estymacja poziomu umiejętności

W badaniu PISA estymacja parametrów, czyli poziomów umiejętności uczniów oraz trudności pytań, odpowiada iteracyjnemu procesowi znalezienia ich najbardziej prawdopodobnych wartości przy zaobserwowanej strukturze odpowiedzi[2].

Ponieważ celem badania jest estymacja wartości poziomu umiejętności w wyróżnionych dziedzinach różnych grup i podgrup uczniów, a nie umiejętności indywidualnych uczniów[3], w analizach stosuje się estymatory *plausible values* (PV). Otrzymujemy je za pomocą procedury

[2] W badaniu PISA estymatorami największej wiarygodności są estymatory WLE (estymator największej wiarygodności Warma).

[3] Estymatory otrzymane poprzez zagregowanie na podstawie estymatorów indywidualnych WLE są estymatorami obciążonymi.

losowania pięciu wartości z rozkładu prawdopodobieństwa wyestymowanego wcześniej poziomu umiejętności. Uśrednienie wyników opartych na wartościach PV daje lepsze oszacowanie wartości parametrów w populacji niż gdybyśmy oparli się na estymatorach WLE (Wu, 2004; OECD, 2005a).

Zadania łączące

W każdej edycji badania PISA jedna z dziedzin jest dziedziną wiodącą. Wyniki z tej dziedziny przeliczane są w taki sposób, że średnia dla krajów OECD wynosi 500, a odchylenie standardowe 100 (w edycji badania, w którym dana skala pojawia się pierwszy raz). Począwszy od roku, kiedy jedna z dziedzin jest wiodącą wyniki z niej na przestrzeni lat są bezpośrednio porównywalne, tak jest: dla *czytania i rozumowania* dla lat badania 2000, 2003 i 2006; matematyki dla lat 2003 i 2006. W przypadku rozumowania w naukach przyrodniczych rok 2006 stanowi podstawę dla przyszłych porównań (OECD, w druku: Annex A, 32).

Takie działanie jest możliwe dzięki zastosowaniu modelu Rascha. Zagwarantowanie baterii wspólnych wskaźników na przestrzeni czasu czy w podgrupach badanych uczniów, pozwala na skalibrowanie względne testów i umieszczenie wszystkich pytań i wszystkich uczniów na jednym kontinuum cechy ukrytej. Ta cecha skali jest nie do przecenienia w przypadku rozbudowanych i wieloaspektowych testów umiejętności.

Literatura

Wu, M. (2004). Plausible Values., *Rasch Measurement Transactions* (18:2): 976-978.

OECD (2005a) *PISA 2003 Data Analysis Manual.* SPSS® Users., OECD, Paris.

OECD (2005b) *PISA 2003 Technical Report.*, OECD, Paris.

OECD (w druku) *PISA Report.*, OECD, Paris.