

Henryk Banaszak Skalowanie (*fragment hasła w Encyklopedii Socjologicznej PWN*)

Do sformułowania problemu skalowania niezbędne są trzy elementy: zbiór **obiektów**, zestaw obserwowalnych cech obiektów, nazywanych **wskaźnikami** oraz założenia na temat **relacji wskazywania**, która wiąże je ze nieobserwowalną **cechą ukrytą**. Skalowanie polega na przyporządkowaniu obiektom wartości cechy ukrytej na podstawie łącznego rozkładu wskaźników. Problemy skalowania rozwiązywane są przez tak zwane **modele skalowania**, czyli algorytmy podejmowania decyzji w procesie wyznaczania wartości ukrytej cechy (cech). Modele skalowania dają się pogrupować ze względu na własności elementów, których użyto w nich do sformułowania problemu skalowania.

Wskaźniki mogą być zmiennymi mierzonymi na „mocnych” (ilorazowa, interwałowa) lub „słabych” skalach pomiarowych (porządkowa, nominalna). Stosunkowo często wskaźniki są zmiennymi dychotomicznymi. O swoistości modelu skalowania stanowi również wbudowane weń założenie na temat poziomu pomiaru cechy ukrytej.

Cechy ukryte mogą przysługiwać obiektom jednego bądź wielu rodzajów. Najczęściej obiektami, którym przypisuje się wartości ukrytej zmiennej są osoby, a czyni się to na podstawie ich reakcji na bodźce reprezentowanych za pomocą zmiennych wskaźnikowych. Istnieją także modele, w których celem skalowania jest przyporządkowanie wartości cechy ukrytej także bodźcom, na które badane osoby reagują¹.

Sposób definiowania relacji wiążącej cechę (-y) ukrytą (-e) może mieć charakter deterministyczny lub probabilistyczny a środki formalne wykorzystywane w takiej definicji ograniczane są przez poziom pomiaru wskaźników i cechy ukrytej.² Sposób definiowania tej relacji jest kluczowym elementem każdego modelu skalowania; w rozbudowanych modelach definicja ta nazywana jest wręcz teorią skalowanej cechy lub teorią reagowania osób w sytuacjach testowych.³

Każdy model skalowania musi zmierzyć się z dwoma zasadniczymi problemami: weryfikacji przyjmowanych w nim założeń oraz uzasadnienia stosowanego w nim sposobu wyznaczania wartości cechy ukrytej na podstawie rejestrowanych wartości wskaźników. W modelach, w których relacja między wskaźnikami i cechą ukrytą jest definiowana w kategoriach probabilistycznych oba problemy przekształcają się w problemy statystyczne: problem testowania hipotezy o zgodności rozkładu wskaźników z założeniami modelu, co zwykle prowadzi do statystycznie interpretowanej miary dopasowania modelu do danych oraz do problemu efektywnej procedury estymacji parametrów charakteryzujących objekty.

Gdy w praktycznej sytuacji skalowania okazuje się, że dany zestaw wskaźników nie spełnia dokładnie założeń modelu, należy rozwiązać dwa kolejne problemy: ustalenia liczby cech ukrytych oraz uzasadnienia kryteriów powiązania z nimi poszczególnych wskaźników. Rozstrzygnięcie pierwszego z nich polega na ustaleniu ile ukrytych cech (nazywanych niekiedy „wymiarami”) dane wskaźniki reprezentują, rozstrzygnięcie drugiego – odpowiedzi na pytanie, które wskaźniki są „reprezentantami” której ze zidentyfikowanych uprzednio cech.

Skalowanie jako samodzielny przedmiot badań metodologów rozwinęło się w drugiej połowie XX wieku. W piśmiennictwie metodologicznym problematyka występuje pod kilkoma nazwami – jako „analiza ukrytej struktury” (Latent Structure Analysis), „analiza ukrytych klas” (Latent Class Analysis, LCA) lub ukrytych cech („traits”, stąd skrót LTA). W piśmiennictwie

¹ Celem testu z języka obcego może być ustalenie poziomu kompetencji językowej jego uczestników lub określenie stopnia trudności poszczególnych pytań testowych bądź jedno i drugie.

² : W analizie ukrytej struktury, która jest szczególnym przypadkiem modelu skalowania zarówno wskaźniki jak i cecha ukryta są zmiennymi nominalnymi. W takim przypadku model nie może zakładać, że wskaźniki są liniowymi funkcjami liniowej cechy ukrytej. Założenie takie jest natomiast założenie to jest w pełni dopuszczalne gdy wskaźniki i cecha ukryta są mierzone na poziomie interwałowym, jak to ma miejsce w analizie czynnikowej, która w swojej eksploracyjnej wersji jest również szczególnym przypadkiem modelu skalowania.

³ Stąd między innymi nazwa Item Response Theory (IRT) nadawana modelom wywodzącym się od skalogramu Rascha

adresowanym bardziej do psychologów niż socjologów nazywana jest „teorią reagowania na bodźce” (Item Response Theory, IRT).

We współczesnej problematyce skalowania występującej w socjologii dają się wyróżnić dwa nurty. Pierwszy, związany z nazwiskiem Guttmana, opiera się na założeniu kumulatywności reakcji, drugi przyjmuje założenie addytywności. W najnowszym piśmiennictwie na temat skalowania, zaawansowanym formalnie, granice między nurtami zaczynają się zacierać.

Model skalowania kumulatywnego (zwany potem skalogramem Guttmana) Louis Guttman (1916-1987) zaproponował w 1950 roku⁴. W modelu tym zakłada się, iż bodźce (pytania) różnią się stopniem „trudności”, zaś osoby, które na nie reagują różnią się stopniem „umiejętności”. Kumulatywność modelu polega na założeniu, iż osoba, której poziom umiejętności pozwala na poprawną reakcję na pewien bodziec reaguje również poprawnie na wszystkie bodźce o niższym stopniu ich „trudności”. Wartość cechy ukrytej w modelu Guttmana (poziom „kompetencji” odpowiadającej osobie) była prostą liczbą „poprawnych” reakcji na bodźce (pytania) zaś stopień „trudności” pytania szacowano za pomocą odsetka osób, które nań poprawnie zareagowały.

Deterministyczny charakter skalogramu Guttmana nie pozwalał na zadowalające rozwiązanie problemu weryfikacji jego założeń oraz uzasadnienie procedury wyznaczania wartości cechy ukrytej. Wad tych nie ma już model kumulatywnego skalowania w probabilistycznej wersji Mokkena⁵. Jego model opiera się na założeniu podwójnej monotoniczności, wedle którego prawdopodobieństwo udzielenia poprawnej odpowiedzi na pojedyncze pytanie jest tym większe, im wyższy jest poziom „umiejętności” odpowiadającego (co pozwala porządkować odpowiadających) a dla każdej pojedynczej osoby prawdopodobieństwo poprawnej odpowiedzi jest wyższe dla pytania „łatwiejszego” niż dla „trudniejszego” (co pozwala z kolei porządkować pytania). Model Mokkena jest nieparametryczny - nie określa postaci funkcji wiążącej prawdopodobieństwo poprawnej odpowiedzi z poziomem „umiejętności” odpowiadającego i ze stopniem „trudności” pytania. Mimo to, jego autorowi i następcom udało się skonstruować efektywne narzędzia do testowania założenia o podwójnej monotoniczności, szacowania (także przedziałowego) poziomu „umiejętności” osób oraz wyznaczyć asymptotyczne rozkłady miernika skalowalności⁶ (z próby). Model Mokkena został uogólniony dla wskaźników mierzonych na skali porządkowej z zachowaniem zdecydowanej większości narzędzi do statystycznego testowania jego własności i estymacji parametrów.

Ideę skalowania Guttmana podjął w parametrycznym ujęciu G. Rasch⁷. Zaproponował on aby na jednej skali mierzyć zarówno poziom „umiejętności” (cechę ukrytą) osoby odpowiadającej na pytanie jak i również wymagający oszacowania parametr charakteryzujący pytanie (stopień jego „trudności”). Prawdopodobieństwo poprawnej odpowiedzi związał z wielkością różnicy między „umiejętnością” i „trudnością” za pomocą krzywej logistycznej. Model Rascha ma charakter kumulatywny, zawiera założenie podwójnej monotoniczności Mokkena, daje się również łatwo uogólnić dla wskaźników politomicznych. Dzięki explicite zdefiniowanej relacji wiążącej wskaźniki z cechą ukrytą dało się zadowalająco rozwiązać oba zasadnicze statystyczne problemy skalowania – estymacji parametrów osób i pytań oraz testowania hipotezy o spełnianiu przez dane empiryczne konsekwencji wynikających z założeń modelu. Rozwinięcia modelu Rascha polegają na uwzględnianiu coraz większej liczby parametrów charakteryzujących osoby oraz wskaźniki. Współczesne prace na ten temat wykazują pokrewieństwa rozbudowanych modeli Rascha i metod analizy czynnikowej, analizy korespondencji oraz analiz typu conjoint.

⁴ Guttman, L. (1950). The basis for scalogram analysis. W S.A. Stouffer i inni [red.] *Measurement and Prediction* (str. 60-90) Princeton, NJ, Princeton University Press

⁵ Mokken, R.J. (1971). *A theory and procedure of scale analysis. With applications in political research*. New York/Berlin: De Gruyter (/Mouton

⁶ Współczynnik skalowalności H pochodzący od Loevingera (Loevinger, (1947) zadający sprawę częstości występowania w próbie odpowiedzi niezgodnych założeniem kumulatywności wykorzystywał także Guttman.

⁷ Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen and Lydiche.

Schemat przedstawiania modelu skalowania

Zakładamy, że badamy losową próbę respondentów za pomocą zestawu wskaźników. Problem skalowania to zestaw założeń i zestaw pytań, na które należy odpowiedzieć, aby zdefiniować model skalowania inaczej zwany skalogramem.

Założenia:

1. Poziom pomiaru wskaźników (binarne, porządkowe, interwałowe)
2. Postulaty nakładane na funkcję reakcji wiążącą zmienną ukrytą (skalę) ze wskaźnikami⁸ (Guttman: kumulatywność deterministyczna, Mokken: kumulatywność probabilistyczna bez zadanej funkcji reakcji, własności funkcji – lokalna niezależność oraz double monotony, Rasch: kumulatywność probabilistyczna z zadaną funkcją reakcji, lokalna niezależność oraz double monotony jako konsekwencja postaci funkcji)⁹
3. Konsekwencje w/w założenia 2 dla łącznego rozkładu wskaźników (rozkładu profili reakcji)¹⁰

Pytania:

1. Jak mierzona jest niezgodność rozkładu empirycznego wskaźników z rozkładem oczekiwanym wedle konsekwencji założeń modelu?
2. Kiedy łączny rozkład wskaźników jest wystarczająco zgodny z konsekwencjami założeń modelu (w szczególności, czy profile reakcji niezgodne z założeniami skali nie występują zbyt często, by założenia odrzucić) - jaka jest reguła podejmowania decyzji w tej sprawie?
3. Jeśli powyższa reguła decyzji ma postać statystyczną, to jaka jest treść hipotezy traktowanej jako zerowa? Jaka jest statystyka testowa? Jaki ma rozkład? Jak wyznacza się obszar krytyczny?
4. Czy stopień zgodności rozkładu empirycznego z konsekwencjami modelu można poprawić:
 - a. Eliminując niektóre wskaźniki
 - b. Przyjmując założenie, że zmienna ukryta (skala) nie jest jednowymiarowa
5. Jeśli tak na 4a, to czy reguła selekcji/eliminacji wskaźników ma charakter reguły wnioskowania statystycznego? Jeśli ma, to jak brzmi hipoteza zerowa, statystyka testowa, jej rozkład oraz jaki jest sposób wyznaczania obszaru krytycznego?
6. Jeśli tak na 4b, to czy reguła podejmowania decyzji o liczbie wymiarów zmiennej ukrytej ma charakter reguły wnioskowania statystycznego? Jeśli ma, to jak brzmi hipoteza zerowa, statystyka testowa, jej rozkład oraz jaki jest sposób wyznaczania obszaru krytycznego?
7. Jaka jest funkcja agregująca profile reakcji w wartości zmiennej ukrytej?
8. Jak jest uzasadniana postać funkcji? Czy opiera się na jakiejś metodzie estymacji?
9. Funkcją jakiej statystyki z próby jest wartość skali-zmiennej ukrytej przyporządkowywana obiektom?
10. Jakie własności pomiarowe ma zmienna ukryta?
 - a. Jakie są dopuszczalne przekształcenia wskaźników?
 - b. Jakie są dopuszczalne przekształcenia statystyk z próby?
 - c. Jaka jest klasa dopuszczalnych przekształceń funkcji agregującej profile reakcji w wartości zmiennej ukrytej.

⁸ Podać zestawy postulatów dla każdego modelu

⁹ Dla nabrania wprawy podać założenia modeli Proctor-a, Goodman-a

¹⁰ Jakie są te konsekwencje dla każdego z modeli?