Henryk Banaszak

Pomiar, skalowanie, model skalowania

Pomiar a skalowanie

Konieczność uwzględnienia elementarnych pojęć *teorii pomiaru* przed prezentacją zagadnień *skalowania* wynika stąd, iż w skalowaniu przyjmuje się założenia na temat pomiarowych właściwości wskaźników zmiennych ukrytych. Poziom pomiaru wskaźników określa z kolei klasę dopuszczalnych ich przekształceń, a w konsekwencji, zestaw parametrów statystycznych, których można używać do opisu rozkładów zmiennych wskaźnikowych. Widać zatem, że poziom pomiaru wskaźników wyznacza modelom skalowania granice, których przekraczanie grozi przyjmowaniem fałszywych przeświadczeń na temat rzeczywistości. Warto zatem mieć na uwadze sposób, w jaki te granice są określane oraz wiedzieć, gdzie przebiegają.

Z powyższego wynika także, że pomiar i skalowanie są dwiema różnymi, rozłącznymi (choć powiązanymi, jak pokażemy dalej) operacjami. W związku z tym bezrefleksyjne zamienne stosowanie obu terminów, tak często niestety spotykane, jest ich drastycznym nadużyciem. Warto zatem pokazać, czym pomiar różni się od skalowania i jakie są relacje między nimi.

Problem pomiaru

Nie zamierzamy tu rekonstruować klasycznej teorii pomiaru, gdyż opis takiej rekonstrukcji można łatwo znaleźć w podstawowych tekstach metodo-

* Henryk Banaszak ukończył socjologię na Uniwersytecie Warszawskim, gdzie się doktoryzował. W Instytucie Socjologii UW prowadzi zajęcia ze statystyki. Niezależny konsultant statystyczny, ekspert MEN i CKE (habanasz@is.uw.edu.pl).

logicznych, a od niedawno także w dobrych podręcznikach statystyki[1]. Przytoczymy natomiast tylko te jej elementy, które bezpośrednio wiążą się z różnicą między pomiarem a skalowaniem.

Pomiar jest operacją przyporządkowywania liczb obiektom empirycznym. Jeśli liczby przyporządkowane obiektom *dobrze reprezentują* ich empiryczne własności, aby opisać całą zbiorowość lub pojedynczy empiryczny obiekt, nie trzeba wykonywać na nich operacji fizycznych, wystarczy wykonać operacje na reprezentujących obiekty liczbach. Właściwa liczbowa reprezentacja empirycznych własności obiektów musi spełniać jeden postulat, nazwijmy go postulatem *odpowiedniości*. Daje się on sformułować tak:

relacje między liczbami przypisanymi obiektom empirycznym powinny **odpowiadać** empirycznym relacjom między obiektami

Dla przykładu, jeśli patyk *a* **wystaje ponad** patyk *b* po ustawieniu obu na stole, czyli stwierdzono **empirycznie**, że *a* jest dłuższy od *b*, to liczba *f(a)* reprezentująca jego długość **musi** być większa od liczby *f(b)* reprezentującej długość patyka *b*, w przeciwnym razie postulat odpowiedniości nie będzie spełniony.

Postulat odpowiedniości ma strukturę prostą lub złożoną, w zależności od tego, jak ubogi lub bogaty jest zestaw empirycznych relacji między obiektami, których własności chcemy reprezentować liczbowo. W związku z tym klasyczna teoria pomiaru zastosowana w odniesieniu do konkretnego zestawu obiektów musi poradzić sobie z dwoma pytaniami:
1. Czy dla danego zestawu empirycznych obiektów i empirycznych operacji istnieje liczbowa reprezentacja spełniająca postulat odpowiedniości?
2. Jeśli istnieje, to do jakiej klasy funkcji należy, to znaczy, jakie jej przekształcenia spełniają postulat odpowiedniości?

Pierwsze z tych pytań nosi nazwę problemu istnienia, pytanie drugi – problemu jednoznaczności funkcji pomiarowej.

Zwróćmy uwagę, jak bardzo „empiryczny” jest sposób definiowania i badania własności funkcji pomiarowych. Sprawdzenie, czy konkretna reprezentacja liczbowa konkretnego zbioru obiektów empirycznych jest pomiarem, ustalenie typu tego pomiaru (jednoznaczności funkcji pomiarowej) jest nie-

[1] Zob. Lissowski, Haman, Jasiński (2008); strony 26–38 oraz 105–108.

wykonalne bez empirycznych operacji na empirycznych obiektach. Pomiar zakłada zatem obserwowalność własności mierzonych. Założenia tego nie przyjmuje się w skalowaniu.

Problem skalowania

Punktem wyjścia problemu skalowania są zmienne zarejestrowane w pewnym zbiorze obiektów, nazywane dalej *wskaźnikami*. Zarejestrowane, oznacza „dostępne empirycznie" lub inaczej „obserwowalne".

Zakłada się, że wskaźniki są wynikiem pomiaru znanego typu, co oznacza, że dla każdego z nich znany jest zakres dopuszczalnych analiz statystycznych, które można na nich wykonywać, a w konsekwencji, repertuar środków formalnych dostępnych przy definiowaniu relacji, w jakie wchodzą one z „cechami ukrytymi".

Cechy ukryte są elementem *teorii zjawiska*, która wiąże obserwacje z konstruktami teoretycznymi za pomocą **reguły korespondencji**. Reguła ta, nazywana skalą, skalogramem lub modelem skalowania *zakłada* pewien zestaw:

- własności wskaźników
- własności cech ukrytych
- relacji (zależności) między cechami obserwowalnymi i ukrytymi

z których *wynikają* reguły wnioskowania o cechach ukrytych na podstawie cech obserwowalnych.

Reguła korespondencji jest elementem **teorii empirycznej**, może być zatem sfalsyfikowana. Problem skalowania, w jego ogólnej postaci, składa się więc z dwóch współzależnych podproblemów:

(1) problemu wykonalności skalowania (oznacza to test teorii, z której wywodzi się reguły korespondencji)

(2) problemu askrypcji, czyli algorytmu przyporządkowywania wartości zmiennej ukrytej obiektów na podstawie wartości wskaźników, które u obiektów zarejestrowano.

Problem (2) pojawia się dopiero po rozstrzygnięciu problemu (1). Wyznaczanie wartości cechy ukrytej dla pojedynczych obiektów ma sens tylko wtedy, gdy reguła korespondencji nie została sfalsyfikowana, to znaczy, gdy własności wskaźników nie przeczą przyjmowanym na ich temat założeniom.

Pomiar a skalowanie

Jak widać z powyższej charakterystyki, mierzenie w naukach społecznych jest aktywnością prateoretyczną albo **ateoretyczną**, choć zarazem ściśle sformalizowaną. Aby *zmierzyć* własności obiektów, aby obiektom przyporządkować liczby reprezentujące te własności, trzeba wiedzieć, **jaki jest empiryczny system**, który te obiekty współtworzą, a następnie ***przeprowadzić dowód*** na istnienie funkcji pomiarowej, która własności te liczbowo wyrazi.

Skalowanie z kolei odwołuje się do teorii empirycznej, z której wynikają reguły korespondencji wiążącej wyniki pomiaru z wartościami cechy ukrytej. Skalowanie *zakłada* zatem istnienie funkcji pomiarowych i znajomość ich własności. Polega najpierw na sprawdzeniu, czy rzeczywistość empiryczna jest zgodna z teorią, a jeśli tak, następnie, na ujawnieniu (przyporządkowaniu) wartości cech ukrytych przysługujących poszczególnym obiektom.

Pomiar i skalowanie mają zatem identyczny cel – przyporządkowanie empirycznym obiektom liczb reprezentujących ich empiryczne własności. Jednak w skalowaniu decydującą rolę w tym przyporządkowaniu odgrywa teoria empiryczna skalowanego zjawiska, która może być prawdziwa lub nie, trafnie lub nietrafnie opisująca badaną rzeczywistość.

W pomiarze problem trafności przyporządkowania liczb obiektom się nie pojawia. W pomiarze, aby odpowiednio przyporządkować liczby obiektom, wystarczy ustalić **formalne własności** empirycznego systemu relacyjnego. Gdy się to wykona, problem istnienia i jednoznaczności liczbowej reprezentacji staje się problemem technicznym. Jeśli własności empirycznego systemu relacyjnego dają się opisać w języku zbiorów i relacji, znalezienie odpowiadającej mu funkcji pomiarowej odbywa się *de facto* automatycznie. Ponadto, każdy empiryczny system relacyjny, mały lub wielki, złożony z rozmaitych empirycznych obiektów (kamyki, patyki, naczynia, owoce, i tak dalej), jeśli ma pewne własności formalne, jest „mierzalny” za pomocą jakiejś funkcji pomiarowej[2].

---

[2] Klasyczna teoria pomiaru zajmuje się właśnie ustalaniem, dla jakich empirycznych systemów relacyjnych, formalnie charakteryzowanych, istnieją liczbowe – i jakiego typu – reprezentacje. Warto zauważyć, że aby stwierdzić, czy dla jakiegoś typu systemu relacyjnego istnieje funkcja pomiarowa jakiegoś rodzaju, teoria pomiaru nie musi zakładać, że w świecie empirycznym istnieje zbiór obiektów posiadający te formalne własności, które istnienie i jednoznaczność funkcji pomiarowej przesądzają. Teoria pomiaru jest

W wypadku skalowania sytuacja jest zasadniczo inna.

Po pierwsze, w skalowaniu własności ukrytych zakłada się wykonanie niebudzącego wątpliwości pomiaru cech obserwowalnych. Zatem ***pomiar poprzedza skalowanie*** i jest warunkiem koniecznym jego wykonalności.

Po drugie, w skalowaniu *zakłada się teorię* obiektów, u których poziomy cechy ukrytej zamierza się zidentyfikować. Teoria ta, jak wspomnieliśmy, może zostać sfalsyfikowana, zatem próba skalowania może zakończyć się niepowodzeniem, gdy teoria empiryczna okaże się sprzeczna z danymi empirycznymi (z wynikami pomiaru wskaźników). Prawdziwość teorii, na której opiera się skalowanie, jest drugim warunkiem jego wykonalności.

Skalowanie jest zatem – *de facto* – operacją testowania jakiegoś fragmentu teorii empirycznej przy użyciu wyników pomiaru zmiennych obserwowalnych w pewnym konkretnym zbiorze obiektów.

Modele skalowania

Modelem skalowania będziemy dalej nazywali zestaw własności elementów niezbędnych do sformułowania problemu skalowania. Własności te dają się pogrupować następująco:

1. **Założenia na temat danych** obejmujące:
   a. poziom pomiaru wskaźników i zmiennej ukrytej;
   b. własności łącznego rozkładu wskaźników i cechy ukrytej, w tym zakładane rozkłady brzegowe zmiennych wskaźnikowych;
   c. własności relacji między wskaźnikami i cechą ukrytą.
2. Własności **algorytmu wyznaczania wartości cechy ukrytej** na podstawie łącznego rozkładu wskaźników; mogą one wynikać z założeń na temat danych lecz nie muszą.
3. Własności **kryteriów oceny jakości skalowania**, zazwyczaj rozumianej jako stopień zgodności danych wejściowych (łącznego rozkładu wskaźników) z rozkładem rekonstruowanym z parametrów modelu.

Istniejące modele skalowania dają się pogrupować ze względu na każdą z powyższych własności. Zanim przedstawimy naszą propozycję klasyfikacji

---

zatem teorią czysto formalną, gdyż o prawdziwości jej twierdzeń nie rozstrzyga doświadczenie, lecz dowód formalny.

metod skalowania, rozważmy repertuar opcji, którymi dysponujemy w każdej z grup własności modelu.

Założenia pomiarowe

Poziom pomiaru wskaźników decyduje przeważnie o kształcie modelu, gdyż określa zakres informacji o obiektach, które można wykorzystać do identyfikowania wartości cechy ukrytej i zarazem przesądza o jej statusie pomiarowym. Rozważmy najpierw przypadek najprostszy, gdy wszystkie zmienne wskaźniki mierzone są za pomocą funkcji pomiarowej tego samego typu[3].

Jeśli wskaźniki mają charakter **nominalny**, bardzo trudno jest uzasadnić tezę o ilościowym charakterze cechy ukrytej odpowiedzialnej za ich łączny obserwowalny rozkład. Cecha ukryta w takiej sytuacji musi być również „mierzona na skali nominalnej”, czyli musi być zwykłą klasyfikacją obiektów. Skalowanie nazywane jest w takich wypadkach *analizą ukrytej struktury* lub *Latent Class Analysis* (LCA).

W naukach społecznych wskaźniki mają najczęściej status **zmiennych porządkowych** i naturalną konsekwencją tego faktu jest przyjmowanie założenia, iż zmienna ukryta ma również porządkowy charakter. Tak na przykład zakłada się w wypadku *skalogramu Mokkena* (patrz dalej w tym tomie). Spotykamy jednak modele, w których mimo porządkowego pomiaru wskaźników cecha ukryta traktowana jest jako zmienna interwałowa, pozwalająca interpretować odległości między obiektami, którym wartości cechy ukrytej przyporządkowano.

Interwałowe wskaźniki są rzadkim luksusem w naukach społecznych, gdyż umożliwiają nie tylko interwałową interpretację wyników skalowania, lecz także pozwalają zdefiniować relację między wskaźnikami a cechą za pomocą funkcji liniowych. Z możliwości tej korzysta się od dawna, czego owocem są wyrafinowane, wielozmiennowe modele skalowania liniowego, takie jak *analiza czynnikowa* czy *modele równań strukturalnych* (patrz dalej w tym tomie).

---

[3] W piśmiennictwie metodologicznym użyto by sformułowania „wskaźniki mierzone są na tej samej skali”.

Założenie o interwałowym charakterze wskaźników nie implikuje interwałowego charakteru zmiennej ukrytej. Istnieje obszerna klasa modeli skalowania, do której należą *modele segmentacyjne*, taksonomiczne lub ogólniej metody klasyfikacji, nazywana także analizą skupień, w których odległości między obiektami przechowywane w interwałowych wskaźnikach wykorzystywane są wyłącznie do dzielenia zbioru obiektów na części. Oznacza to, iż cecha ukryta jest w tych modelach zmienną nominalną.

Specyficzną kategorię modeli skalowania tworzą modele, w których wskaźniki są zmiennymi dwuwartościowymi, zwykle **binarnymi**. Wyjątkowy status pomiarowy zmiennych binarnych[4] i ich prostota zaowocowały powstaniem bogatej rodziny modeli *skalowania kumulatywnego*, które wraz z współczesnymi, rozwiniętymi modelami skalowania umiejętności za pomocą testów współtworzą *Item Response Theory* (IRT). Wspólnym rdzeniem modeli z grupy IRT jest teoria reagowania na pojedyncze pytania, w której wynik odpowiadania zależy zarówno od poziomu cechy ukrytej (umiejętności) odpowiadającego, jak i własności (trudności) pytania (patrz dalej w tym tomie).

Oddzielną kategorię modeli skalowania stanowią te, w których wskaźniki mierzone są na różnych poziomach pomiaru. Mieszanina nominalnych, porządkowych i interwałowych wskaźników stawia szczególne i wyjątkowo wysokie wymagania przed teorią empiryczną, z której wywodzi się związek wskaźników z cechą ukrytą. W niniejszym tomie nie będziemy się zajmować tego typu modelami skalowania.

Relacja wskaźniki – cecha ukryta

Ze względu na sposób definiowania relacji między zmienną ukrytą i jej obserwowalnymi wskaźnikami, modele skalowania dzielą się na dwie klasy:

- modele deterministyczne, w których zakłada się funkcyjną zależność cechy ukrytej od wskaźników lub wskaźników od cechy ukrytej.
- modele probabilistyczne, w których związek ten ma charakter po części losowy.

---

[4] Z formalnego punktu widzenia zmienna binarna mierzona jest za pomocą funkcji pomiarowej wyznaczanej z dokładnością do przekształcenia liniowego, co oznacza, że zmienne binarne można bez utraty informacji o obiektach przekształcać liniowo.

Tabela 1. Zestawienie popularnych metod analizy danych traktowanych jako szczególne przypadki modeli skalowania

Model	**Poziom pomiaru wskaźników**	**Rodzaj zależności wskaźników od cech ukrytych**	**Poziom pomiaru cechy ukrytej**
Analiza ukrytej struktury Lazarsfelda	Nominalny Binarny	Probabilistyczny	Nominalny
Analiza skupień K-Means	Interwałowy	Deterministyczy	Nominalny
Probabilistyczne metody analizy skupień	Nominalny Binarny Interwałowy	Probabilistyczny	Nominalny
Skalogram Guttmana	Binarny	Deterministyczy	Porządkowy
Skalogram Mokkena	Binarny Porządkowy	Probabilistyczny	Porządkowy
Skalogram Rascha	Binarny Porządkowy	Probabilistyczny	Interwałowy
Eksploracyjna analiza czynnikowa	Interwałowy	Probabilistyczny	Interwałowy
Model równań strukturalnych	Interwałowy	Probabilistyczny	Interwałowy

Deterministyczny charakter miały najstarsze wersje skalowania kumulatywnego z pierwszej połowy ubiegłego wieku (skala uprzedzeń etnicznych Bogardusa[5], a także skala stresu wprowadzona w IV tomie *American Soldier* przez Guttmana[6]). Również za deterministyczny należy uznać związek, jaki między wskaźnikami a cechą ukrytą zakładają *de facto* nieprobabilistyczne modele segmentacyjne jak na przykład K-Means.

Deterministyczny charakter związku wskaźników z cechą skalowaną jest istotną przeszkodą w rozwiązywaniu podstawowych problemów, które w praktyce skalowania trzeba rozwiązywać. Znaczna część z nich staje się, jak pokażemy dalej, rozwiązywalna po założeniu probabilistycznego charak-

[5] E.S Bogardus. *A Social Distance Scale*, „Sociology and Social Research”, 1933, vol. 17, s. 265–271.

[6] L. Guttman 1950. *The basis for scalogram analysis*, w: Stouffer et al., *Measurement and Prediction. The American Soldier*, Vol. IV. New York: Wiley.

teru tego związku. Zanim do tych problemów przejdziemy, spójrzmy jak przy użyciu obu kryteriów daje się scharakteryzować popularne w naukach społecznych metody i techniki analizy danych, jeśli potraktować je jako szczególne przypadki skalowania.

Oprócz tego modele skalowania dają się podzielić ze względu na sposób definiowania własności wskaźników i ich związku wartościami cechy ukrytej, na modele kumulatywne i addytywne. W modelach kumulatywnych (Guttmana, Mokkena, Rascha) wskaźniki są porządkowane ze względu na „trudność”, jaką stanowią dla (odpowiadających) obiektów, a każdemu poziomowi trudności odpowiada poziom cechy ukrytej (jak w skoku wzwyż).W modelach addytywnych (zwłaszcza liniowych) założenie to nie występuje.

Podstawowe problemy skalowania

Proces skalowania rozpoczyna się od utworzenia łącznego rozkładu zmiennych wskaźnikowych i kończy (ewentualnie) przyporządkowaniem wartości cech ukrytych obiektom zbioru, w którym zmienne wskaźnikowe są zdefiniowane. Jego uwieńczeniem jest opis własności cechy ukrytej, jej rozkładu i relacji ze wskaźnikami, a także ocena jakości skalowania. Ujmowany pragmatycznie proces skalowania jest sekwencją problemów decyzyjnych, które muszą być rozstrzygnięte, aby skalowanie przeprowadzić.

Przedstawimy teraz listę tych problemów, z którymi, naszym zdaniem, musi sobie poradzić każdy model skalowania.

I. Problem skalowalności

Jak wspomnieliśmy poprzednio, skalowanie jest *de facto* testowaniem fragmentu teorii obiektów, które pragnie się opisać za pomocą cechy ukrytej. Nie ma zatem powodu, aby zakładać *a priori*, że fragment teorii wykorzystywany w modelu skalowania jest zgodny z rzeczywistością. Posiadanie miary stopnia zgodności założeń modelu (teorii) z własnościami empirycznego rozkładu zmiennych obserwowalnych jest zatem niezbędnym elementem dobrego modelu skalowania, bowiem jeśli model takiej miary nie przewiduje, nie sposób odpowiedzieć na pierwsze dwa oczywiste pytania (1) i (2):

(1) Jak dalece łączny rozkład wskaźników jest zgodny z modelem? Jak dobrze model pozwala odtwarzać łączny rozkład wskaźników?
(2) Czy zbiór wskaźników jest skalowalny, to znaczy, czy stopień zgodności danych z modelem jest wystarczający?

II. Problem liczby wymiarów cechy ukrytej i relacji między nimi

W dotychczasowych rozważaniach milcząco założyliśmy, że na podstawie jednego zestawu wskaźników można próbować ustalać wartości jednej tylko cechy ukrytej. Pora uchylić to założenie, gdyż nie zawsze ma ono teoretyczne podstawy. Zbyt mała liczba wymiarów cechy ukrytej jest czasami (na przykład w analizie czynnikowej, a także w skalowaniu kumulatywnym) najważniejszym źródłem niezgodności przewidywań modelu z danymi empirycznymi. Dobry model skalowania powinien w takich sytuacjach umożliwiać sprawdzenie hipotezy, iż w badanej zbiorowości cecha ukryta jest więcej niż jednowymiarowa, a jeśli hipoteza taka zostanie potwierdzona, powinien pokazywać, w jakich relacjach te wymiary względem siebie pozostają, czy zasadne jest założenie o ich niezależności czy też nie. Powinien zatem odpowiadać na pytania (3) i (4):

(3) Ile cech ukrytych (wymiarów zmiennej ukrytej) trzeba założyć, aby dany zbiór wskaźników (w danym zbiorze obiektów) był skalowalny?
(4) W jakich relacja pozostają względem siebie wymiary cechy ukrytej?

III. Czy wszystkie wskaźniki są potrzebne?

Nietrafność przewidywań modelu może wynikać nie tylko z niedostatecznej liczby wymiarów cechy ukrytej uwzględnianych przez model, ale także z nietrafnego doboru wskaźników. Sytuacja taka zdarza się często przy bezrefleksyjnym stosowaniu modeli skalowania w międzynarodowych badaniach porównawczych, gdy wskaźniki efektywne w diagnozowaniu cechy ukrytej w jednym kraju okazują się zupełnie nieskuteczne w innym. Dobry model skalowania powinien takie sytuacje wykrywać i umożliwiać uzasadnienie odpowiedzi na pytania (5) i (6):

(5) W jakich relacjach pozostają poszczególne wskaźniki z poszczególnymi wymiarami cechy ukrytej?
(6) Czy w zbiorze wskaźników są pozycje zbędne? Czy są wskaźniki (pozycje testu), z których bez szkody dla skalowania można zrezygnować?

IV. Jakie są własności diagnostyczne poszczególnych wskaźników?

Nie zawsze jedynym celem skalowania jest przyporządkowanie obiektom wartości cechy ukrytej. W badaniach edukacyjnych równie ważnym zadaniem skalowania jest dostarczenie informacji o pojedynczych pytaniach testu, o stopniu ich trudności, wrażliwości na niskie i wysokie poziomy wartości cechy ukrytej, reszcie, o ich rzetelności. Dlatego dobry model skalowania powinien umożliwiać wyznaczanie parametrów pojedynczych wskaźników i odpowiadać na pytanie (7):

(7) Jakie są parametry wskaźników?

V. Jak skalować

W prostych, deterministycznych modelach skalowania wyznaczanie wartości cechy ukrytej na podstawie profilu reakcji na pytania wskaźnikowe jest zabiegiem oczywistym i niepodlegającym ani teoretycznej refleksji ani empirycznej weryfikacji trafności. W pierwszych deterministycznych modelach kumulatywnych, a także w początkach tak zwanego „skalogramu Likerta" wartości cechy ukrytej wyznaczano zliczając poprawne odpowiedzi na pytania rozstrzygnięcia lub dodając do siebie rangi nadawane przez respondentów opiniom przedstawianych im do oceny. W rozwiniętych probabilistycznych modelach skalowania, niekoniecznie parametrycznych (jak na przykład skalogram Mokkena) reguła wyznaczania wartości cechy ukrytej wynika z jego założeń. W niektórych sytuacjach reguła ta nie musi być jednoznaczna (na przykład w modelu Lazarsfelda), wówczas musi jednak umożliwiać wyznaczenie rozkładu cechy ukrytej i uzasadnienie jego wyboru. Musi zatem odpowiadać na pytania (8) i (9):

(8) Jak przyporządkować obiektom wartości zmiennej ukrytej?
(9) Jaki rozkład ma cecha ukryta w danym zbiorze obiektów?

Dobry model skalowania

Dobry model skalowania to taki, który dostarcza trafnej i rzetelnej informacji o poziomie i rozkładzie cechy ukrytej w zbiorze obiektów charakteryzowanych przez zestaw wskaźników o znanych własnościach pomiarowych. Zatem dobry model skalowania:

(a) gwarantuje niezmienniczości wyników skalowania przy dopuszczalnych poziomem pomiaru przekształceniach wskaźników;

(b) gwarantuje optymalności algorytmu skalowania;

(c) gwarantuje jednoznaczność i przekonujące uzasadnienia dla decyzji, które trzeba podejmować rozwiązując problemy (1) – (9) wymienione wyżej.

W następnych rozdziałach zastosujemy powyższe postulaty do oceny kolejnych wersji modeli skalowania kumulatywnego, Zostaną one także zastosowane do oceny metod segmentacji (analizy skupień) traktowanej jako szczególna wersja modelu skalowania z nominalną cechą ukrytą (patrz tabela 1 powyżej). W obu przypadkach okaże się, że aby model skalowania powyższe postulaty spełniał, musi zakładać probabilistyczny charakter związku obserwowalnych reakcji (wskaźników) ze zmienną ukrytą, gdyż dzięki temu podstawowe decyzyjne problemy skalowania stają się statystycznymi problemami estymacji lub testowania hipotez. Nie oznacza to, że są to zawsze problemy rozstrzygalne lub rozstrzygalne łatwo. Jeśli jednak są, oferują odpowiedzi dobrze uzasadnione.