

Własności cech diagnostycznych w badaniach typu taksonomicznego

Properties of the Variables in Taxonomic Type of Research

Danuta Tarka

Politechnika Białostocka, Wydział Zarządzania, Katedra Informatyki Gospodarczej
i Logistyki, email: dtarka@pb.edu.pl

Abstract

Choice of the variables (features) is the one of the most important steps in taxonomic type of the research. Variables used in applied research should be selected carefully. Author presents and discusses in this paper the problem of criteria diagnostic variables should satisfy to give proper results.

Keywords: *taxonomic research, diagnostics variables, choice of variables*

Badanie typu taksonomicznego

Zanim przystąpimy do omawiania problemu cech diagnostycznych, wyjaśnimy, co rozumiemy przez określenie „typu taksonomicznego”. Historycznie, taksonomia, jako dyscyplina naukowa wywodzi się z nauk przyrodniczych. Samo słowo jest złożeniem dwóch słów greckich: *taksis* – (układ, porządek) i *nomos* – prawo, zasada. Jest to, więc dyscyplina naukowa tworząca zasady porządkowania, czyli klasyfikacji¹. Wprowadzenie do taksonomii w ujęcia ilościowego przypisuje się niemieckiemu przyrodnikowi F. Heinckemu oraz polskiemu antropologowi J. Czekanowskiemu. Rozwój procedur taksonomicznych początkowo miał miejsce, przede wszystkim, w naukach przyrodniczych, antropologii, psychologii, później wprowadzono ich stosowanie do nauk geograficznych, ekonomicznych, technicz-

¹ Często można spotkać w literaturze społeczno-ekonomicznej jako synonimy klasyfikacji pojęcia: typologia, delimitacja, grupowanie, rejonizacja, regionalizacja; patrz np. Firich (1957), Fajferek (1966), Gorzelak (1975), Kosiedowski i in.(1989), Chojnicki i Czyż (2006).

nych i in. W wyniku rozwoju procedur taksonomicznych na gruncie wielu dziedzin nauki powstała „(...) bogata terminologia i różne definicje (nawet tych samych terminów)”², co zaowocowało dużą niejednorodnością terminologiczną, nawet w rozumieniu pojęcia taksonomia. Współcześnie, znalezienie monografii, czy artykułu z pojęciem taksonomia w tytule okazuje się trudne. Pojęcie to, jako nazwa zbioru technik (metod) analizy prawie zniknęło z literatury, zarówno polskiej³, jak i anglojęzycznej. Jak stwierdza Balicki (2009), pojęcie taksonomia numeryczna jest przede wszystkim stosowane w naukach biologicznych, w psychologii używa się pojęcia Q-analiza, w naukach technicznych - a zwłaszcza w metodach sztucznej inteligencji – używa się pojęcia rozpoznawanie obrazów⁴. Nawet na gruncie jednej dyscypliny – ekonomii – także nie ma zgody, co do tego, co to jest taksonomia, a co za tym idzie jak używać tej nazwy, często autorzy operują pojęciem analiza skupień, czy klasyfikacja jako synonimami taksonomii⁵. Dla niektórych jednak, taksonomia to coś więcej niż analiza skupień, a jednocześnie coś innego niż klasyfikacja⁶. Wielu też autorów operuje pojęciem metod statystycznej analizy wielowymiarowej⁷ (SAW⁸) czy statystycznej analizy porównawczej⁹ (SAP), do których zaliczają taksonomię¹⁰, jako jeden z jej „działów” lub utożsamiają te pojęcia. W literaturze polskiej wprowadzono też pojęcie wielowymiarowej analizy porównawczej¹¹ (WAP) jako zbioru metod służących konstrukcji miar syntetycznych oraz liniowego porządkowania z ich pomocą obiektów wielocechowych¹².

Rozwój metod obliczeniowych spowodował nie tylko wzrost zastosowań klasycznych metod taksonomicznych, ale także wzrost ilości procedur z zakresu szeroko rozumianych metod klasyfikacji i analizy danych, które to pojęcie jest coraz szerzej spotykane zamiast kategorii metody taksonomiczne. Coraz częściej można

² Balicki A. (2009, s. 208).

³ Większość monografii z taksonomią w tytule pochodzi z okresu do początku lat dziewięćdziesiątych; Pocięcha i in. (1988), Grabiński i in. (1989), Nowak (1990), Grabiński (1991).

⁴ Także wprowadzone do nauk ekonomicznych, patrz np. Rozin (1979), Jajuga (1990).

⁵ Np. Balicki (2006, s. 204 i nast.), dyskusję na temat rozumienia pojęcia taksonomia patrz także Pocięcha i in. (1988).

⁶ Patrz np. Marek i Noworol (1987, s. 184).

⁷ Np. wg Kukuły (2000, s. 14) J.Czekanowski był w Polsce prekursorem statystyki wielowymiarowej a nie taksonomii.

⁸ np. Jajuga (1993), Balicki (2009).

⁹ np. Panek (2008), dla którego metody porządkowania obiektów są odrębnym zbiorem procedur od metod grupowania. Nie używa pojęć taksonomia czy klasyfikacja.

¹⁰ Patrz np. Nowak (1990); utożsamia on taksonomię z klasyfikacją (s. 15).

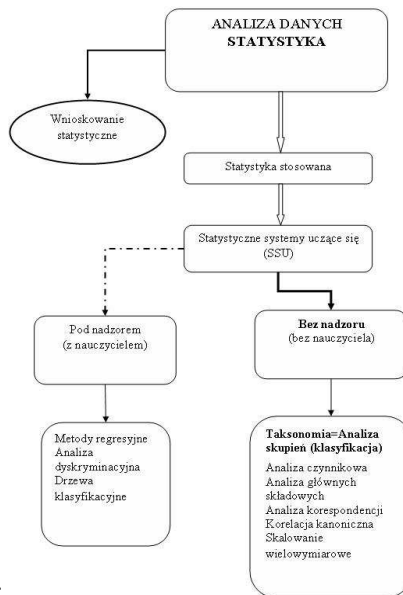
¹¹ Prekursorem WAP był Z.Hellwig (1968).

¹² Patrz Grabiński (1988, s. 239; 2003, s. 5), gdzie wyraźnie rozróżnia WAP i taksonomię numeryczną, z kolei np. Panek (2009) do metod WAP zalicza zarówno metody porządkowania liniowego i nieliniowego (czyli grupowanie), metody analizy czynnikowej, korelację kanoniczną, analizę korespondencji jak i analizę dyskryminacyjną.

spotkać się z pojęciem klasyfikacji, czy grupowania w monografiach o „systemach uczących się”¹³.

Podsumujmy więc, gdzie współcześnie znajduje się taksonomia na mapie nauki jako zbiór metod (procedur) klasyfikacji, w klasycznym tego słowa znaczeniu. Taksonomia jest najczęściej synonimem pojęcia klasyfikacja¹⁴ (classification) lub analiza skupień (cluster analysis)¹⁵ i w takim znaczeniu jest rozumiana w tej pracy, chociaż z powodu pewnych różnic zakresowych użyto w tytule określenia „typu taksonomicznego”. W pracach mających na celu klasyfikacje obiektów, badacze często używają analizy czynnikowej i metody głównych składowych, określając badanie jako taksonomiczne, patrz np. Parysek (1982), stąd rozszerzenie pojęcia.

Miejsce taksonomii wśród innych metod analizy danych pokazuje rys. 1.



Źródło: opracowanie własne.

Rys. 1. Taksonomia na tle innych metod analizy danych statystycznych

¹³ Patrz np. Koronacki i Ćwik (2005), Krzyśko i in. (2008).

¹⁴ Pocięcha (2008, s. 1) „Można więc najogólniej twierdzić, że taksonomia jest nauką o zasadach klasyfikacji, a wręcz synonimem słowa klasyfikacja.”

¹⁵ Choć bardziej poprawne będzie stwierdzenie, że klasyfikacja i/lub grupowanie są synonimami pojęcia taksonomia, które zaczęło wychodzić „z mody”, jak się wydaje. Są także autorzy, dla których klasyfikacja jest pojęciem szerszym niż grupowanie. Do metod klasyfikacji zaliczają także techniki analizy czynnikowej.

Pojęcie „klasyfikacja” ma wiele znaczeń i w niniejszej pracy do omówienia problemu cech diagnostycznych będziemy mieli na myśli pierwsze znaczenie wg poniższego rozróżnienia T. Borysa (1984). Wyróżnia on trzy podstawowe znaczenia pojęcia klasyfikacja:

1. jako czynność polegającą na podziale zbioru obiektów na rozłączne i wyczerpujące podzbiory (grupy, skupiska), z punktu widzenia kryterium badania i w oparciu o określone atrybuty obiektów,
2. jako zbiór reguł, na podstawie których dokonywany jest podział,
3. jako wynik podziału zbioru obiektów.

Klasyfikację zaczyna się od określenia celu badania, który z kolei określa zakres badania i dobór atrybutów (cech¹⁶) opisujących klasyfikowane obiekty, stąd w niniejszej pracy interesuje nas tylko pierwsze znaczenie słowa klasyfikacja. Cel i dobór cech są etapami wstępnymi przy każdym badaniu, mającym za zadanie klasyfikację obiektów, czy to typowymi metodami grupowania, czy z użyciem analizy czynnikowej i technik pochodnych.

Własności cechy diagnostycznych

Jak zauważa T. Panek (1980) „(...) zagadnienie prawidłowego doboru zmiennych, pomimo, że jest jednym z najważniejszych czynników decydujących o ostatecznej wartości analizy, wydaje się nie w pełni docenianym w praktyce badań statystycznych. Chociaż większość autorów zgadza się z powyższą tezą, rzadko daje temu wyraz w sposobie prowadzenia przez nich prac, poświęcając z reguły stosunkowo mniej uwagi doborowi zmiennych niż innym etapom badań.” Powyższe słowa pochodzą sprzed wielu lat, ale nadal nie straciły na aktualności. Część autorów w ogóle nie porusza problemu doboru cech diagnostycznych w swych badaniach, część zaś stwierdza co prawda, że jest to istotny problem, ale nie poświęca mu zbyt wiele miejsca. Jak istotny jest to problem pokazano m.in. w pracach Tymowska (1979), Rutkowski (1982), Kowerski (1983) omawiających istotne błędy popełnione przez badaczy i ich wpływ na wyniki badania.

W niniejszej pracy dokonamy krótkiego przeglądu propozycji, zasadniczo, trzech autorów, którzy szerzej zajmowali się problemem własności, jakie powinny mieć zmienne diagnostyczne. Większość bowiem autorów odwołuje się, mniej lub bardziej explicite, do tych propozycji.

¹⁶ W literaturze dotyczącej systemów uczących się używa się pojęcia atrybutu lub zmiennej, w literaturze społeczno-ekonomicznej używa się pojęcia cechy lub zmiennej. W niniejszej pracy, nie wdając się w dyskusje nad różnicami semantycznymi tych pojęć, będziemy używać pojęcia cecha diagnostyczna i zmienna zamiennie.

Etap doboru cech można rozłożyć na dwie części: - wspólną dla wszystkich procedur badawczych analizę własności, które powinny mieć cechy uznane za diagnostyczne oraz – analizę indywidualną dla każdego badania, zależną od typu badanego procesu i metody doboru cech. Cechy powinny mieć bowiem określone własności związane z typem procedury użytej do analizy¹⁷.

Uznaje się¹⁸, że jako pierwszy w Polsce metody taksonomiczne wprowadził do nauk społeczno-ekonomicznych J. Fierich w pracy z 1957 r. W tejże pracy poruszył też, jako pierwszy, problem zasad doboru cech diagnostycznych. Określił on (Fierich 1957) poniższy zestaw postulatów określających, jakie własności powinny mieć cechy diagnostyczne¹⁹ służące klasyfikacji obiektów:

1. ujmować najbardziej istotne właściwości analizowanych zjawisk, zatem do badania należy brać jedynie cechy niezbędne,
2. być prosto i logicznie ze sobą powiązane,
3. być jasno i ściśle sprecyzowane oraz bezpośrednio lub pośrednio mierzalne oraz wyrażalne za pomocą wielkości bezwzględnych (podobieństwo wielkości) lub względnych (podobieństwo struktury),
4. posiadać wysoką zmienność przestrzenną w zbiorze badanych obiektów i niełatwo ulegać wpływom środowiska,
5. być niezależnymi od siebie, ale związanymi z nie uwzględnionymi w badaniu cechami,
6. posiadać wysokie współczynniki zmienności w ramach zbiorowości wyjściowej, zaś w obrębie grup – możliwie niskie.

Powyższe własności cech diagnostycznych są podawane do dziś jako obowiązujące w polskiej literaturze. Kolejni autorzy często podają powyższe własności w inaczej sformułowanych postaciach, rozbitych na ich większą ilość bądź skoma-

¹⁷ Przykładem może być analiza czynnikowa i metoda głównych składowych, użycie których wymaga spełnienia pewnych założeń formalnych (matematycznych) by można było je zastosować.

¹⁸ Jak pisze J. Pocięcha: „Pierwszego zastosowania metody Czekanowskiego w badaniach społeczno-ekonomicznych dokonał Jerzy Fierich (1900-1965), założyciel (1950) i pierwszy kierownik Katedry Statystyki na ówczesnej Wyższej Szkole Ekonomicznej w Krakowie. Podczas II wojny światowej, prof. Czekanowski, jako wysiedleniec, spędził pewien czas w majątku prof. Fiericha w Broniszowie koło Ropczyc. Tam z pewnością dyskutowali o taksonomicznej metodzie Czekanowskiego. Profesor Fierich powrócił do metody Czekanowskiego w publikacji (Fierich J. (1957). Praca ta stała się początkiem szeregu badań taksonomicznych, zrealizowanych w ośrodku statystycznym WSE a następnie AE w Krakowie, bazującym na dorobku Czekanowskiego i Fiericha.” (J. Pocięcha 2008, s.1). Szerzej na temat początków metod taksonomicznych w Polsce patrz cytowana tu publikacja J. Pocięchy z okazji 90-lecia GUS (2008) dostępna na stronie internetowej GUS, patrz także np. Kosiedowski i in. (1989).

¹⁹ Podane za pracą Podolec i Zajac (1978, s. 25 i nast.), patrz też Pocięcha J. i in. (1988, s.19)

sowanych²⁰. Zgodnie z tym, co zostało powiedziane powyżej, określmy, które z własności muszą być spełnione we wszystkich badaniach, a które są specyficzne, czyli związane z procedurami badawczymi. Fierich tego nie rozróżnia. Do przeprowadzenia badania typu klasyfikacja obiektów, wszyscy późniejsi autorzy z tym się zgadzają, należy użyć cech o własnościach 1-3. Kwalifikację pozostałych własności omówimy po prezentacji propozycji następnego autora.

Według Gorzelaka (1979) dobry zestaw zmiennych to zmienne, które:

1. są ściśle zdefiniowane,
2. są jednoznaczne,
3. możliwie dokładnie reprezentują zjawiska i procesy wchodzące w zakres badania,
4. mają dużą zawartość informacyjną (jest to konsekwencja dążenia do uwzględnienia możliwie małej ilości zmiennych – jak precyzuje autor)
5. są nieskorelowane.

Jak widać, powyższy zestaw własności cech wg Gorzelaka, niczym nie różni się zasadniczo od zestawu Fiericha. Pierwsza i druga własność wg Gorzelaka powielają się, ściśle zdefiniowanie powinno być jednoznaczne²¹ – jest to inaczej zapisany postulat trzeci w zestawie Fiericha. Nieskorelowanie odpowiada postulatowi szóstemu, dużą zawartość informacyjną²² można uznać za odpowiednik postulatu pierwszego u Fiericha.

Gorzelak wprowadza także podział własności cech²³ na własności *merytoryczne* i *formalne*. Jako *merytoryczne* określa własności 1,2 i 4 (czyli u Fiericha są to punkty 1-3). Pozostałe postulaty zalicza do kryteriów (własności) formalnych. Nie

²⁰ Autorzy często używają też pojęcia kryterium doboru cech (zmiennych) a nie własność. Tak postępuje np. Grabiński w swych pracach (Grabiński 1988, 1991, 1992). Jest to autor najszerszej dyskutujący nad tym problemem. Czy można zastąpić pojęcie własności pojęciem kryterium? Wydaje się, że nie. Własność jest to w mniemaniu autorki pewien atrybut cechy, niekoniecznie mierzalny. Kryterium diagnostyczności cechy sugeruje możliwość precyzyjnego pomiaru, możliwość dokładnego określenia czy cecha spełnia to kryterium.

²¹ Fierich precyzował własności cech w bardziej ogólnym kontekście badania klasyfikacyjnego, gdy tymczasem Gorzelak analizuje cechy pod węższym kątem, w zastosowaniach do rangowania obiektów i stąd jego wymóg jednoznaczności. Ale pojęcie to dotyczy bardziej kryterium oceny cechy w kontekście: stymulanta versus destymulanta. Bliżej patrz Gorzelak (1979, s. 19).

²² Gorzelak postuluje traktowanie zmiennych jako zbiorów informacji, stwierdzając „Poszczególne zmienne pokrywają swoim zasięgiem pewne informacje o obiektach. Możemy więc traktować zmienne jako zbiory informacji, będące podzbiórami ogólnej zbiorowości informacji, składających się na naszą wiedzę o obiektach. Zagadnienie doboru zmiennych można w takim ujęciu przedstawić jako zagadnienie klasyfikacji informacji.” Gorzelak (1979, s. 19), klasyfikacja zaś by była poprawna musi spełniać pewne kryteria.

²³ Używa przy tym słowa kryterium a nie własność, choć z kontekstu wynika, że traktuje je zamienne, Gorzelak (1979, s. 20), patrz też przypis poprzedni.

definiuje tych pojęć, ale z kontekstu wynika wyraźnie, że merytoryczne własności są związane z celem i dziedziną badania. Do powyższego zestawu cech-kryteriów dodaje, za innymi autorami²⁴, postulaty *formalne*: mierzalność, co najmniej w skali rangowej (u Fiericha własność 3), (6) ciągłość²⁵ zmiennych²⁶, (7) dodatniość związku z konstruowaną zmienną wyjścia, (8) liniowość związków pomiędzy zmiennymi²⁷. Pozostałe wymagania określa jako ściśle wiążące się z metodami doboru zmiennych. Są to: (9) zdolność do dyskryminacji obiektów (u Fiericha własność 4), (10) zgodność proporcji między liczbą zmiennych charakteryzujących zjawiska z proporcjami istotności tych zjawisk (wiąże się to ściśle z problemem doboru wag oraz konstrukcją miary syntetycznej), (11) możliwość wzajemnej kontroli²⁸ oraz (12) reprezentatywność zmiennych pozostawionych względem zmienionych wyeliminowanych (u Fiericha własność 5). Widać tu znów wyraźny wpływ postulatów Fiericha.

Ostatni z omawianych autorów, T. Grabiński, omawiał problem własności cech diagnostycznych w kilku pracach. Poniżej przytoczymy jego postulaty za pracą Grabiński (1988)²⁹, a następnie podsumujemy rozważania jego oraz pozostałych autorów.

Grabiński stwierdza, że zmienne diagnostyczne powinny³⁰:

1. ujmować najbardziej istotne właściwości analizowanych zjawisk i dokładnie je reprezentować (F1),
2. być prosto, jasno i ściśle sprecyzowane (F2),
3. być logicznie ze sobą powiązane (F2),
4. zawierać duży ładunek informacji (F1,G4)
5. być bezpośrednio lub pośrednio mierzalne, co sprowadza się do istnienia wiarygodnych i łatwo osiągalnych danych statystycznych (F3),

²⁴ Pluta (1972), Cieślak (1974), Polak i Starzewska (1975).

²⁵ Rozumianej matematycznie, czyli dziedziną cech powinny być liczby rzeczywiste.

²⁶ Wraz z rozwojem metod „typu taksonomicznego” warunek ten zniknął jako zbyt ograniczający możliwości analizy. Patrz przypis nr 36.

²⁷ Dodatniość i liniowość związków dotyczą tylko badań mających na celu ranking obiektów oparty o miarę syntetyczną.

²⁸ „Oznacza to znajomość związków - statystycznych i merytorycznych – zachodzących pomiędzy poszczególnymi zmiennymi.” Gorzelak (1979, s. 21).

²⁹ Patrz prace Grabiński (1988, s.239-241), (1991, s.43-44), (1992, s.29-30), przy czym sam autor nie jest konsekwentny podając poszczególne kryteria (słowo to stosuje jako synonim własności) w różnych konfiguracjach i nie zawsze identycznym zapisie; dotyczy to zwłaszcza kryteriów formalnych i statystycznych.

³⁰ W nawiasach podano, której własności wg Fiericha (np. F1) lub Gorzelaka (np. G1) odpowiada dany postulat Grabińskiego.

6. być wyrażone w jednostkach naturalnych, a nie wartościowych, raczej w postaci wskaźników natężenia niż w wielkościach absolutnych,
7. charakteryzować się wysoką zmiennością przestrzenną (postulat dużego stopnia informacyjności) (F4),
8. nie być wzajemnie wysoko skorelowane (postulat braku redundancji) (F5),
9. być wysoko skorelowane ze zmiennymi niediagnostycznymi oraz zmienną syntetyczną (F5),
10. umożliwiać wzajemną kontrolę (poprzez znajomość związków statystycznych i merytorycznych między poszczególnymi zmiennymi),
11. nie opisywać zjawisk i procesów specyficznych,
12. charakteryzować się zgodnością proporcji, między liczbą zmiennych charakteryzujących dany aspekt badanych zjawisk a ich znaczeniem merytorycznym.

Porównując powyższy zestaw wymogów z proponowanymi wcześniej przez Fiericha i Gorzelaka, można stwierdzić, że opiera się zasadniczo na pracach poprzedników. Większość jego postulatów pokrywa się z propozycjami, przede wszystkim Fiericha oraz Gorzelaka. Do postulatów poprzedników dodaje dwa, szósty i jedenasty. Grabiński proponuje też bardziej szczegółowy podział kryteriów-własności cech dodając trzeci typ – kryterium statystyczne.

Kryterium merytoryczne określa własności cech związane z dziedziną, zakresem i rodzajem badania. Nie one zależą od techniki badawczej. Są więc najbardziej ogólnymi własnościami atrybutów badanych obiektów bez względu na dyscyplinę, w której stawiane są cele i pytania badawcze. Są to postulowane przez Fiericha własności 1-3, także u Gorzelaka 1-3. Według Grabińskiego są to własności 1-3, 10 i 12. Jak widać z tych trzech podstawowych w literaturze polskiej³¹ analiz, zasadniczo nadal obowiązują postulaty Fiericha jako podstawowe, co do tego istnieje zgodność, choć postulaty te często są sformułowane nieco innymi słowami. Grabiński określił własności dziesiątą i dwunastą jako merytoryczne³², ale wydają się one wątpliwe, zwłaszcza dwunasta zdecydowanie nie jest merytoryczna. Przypomnijmy, że omawiany jest tu problem własności cech diagnostycz-

³¹ Jak stwierdza Gorzelak (1979, s. 20), literaturze anglojęzycznej dotyczącej metod taksonomicznych brak „(...) zainteresowania doбором zmiennych w klasycznej teorii taksonomii czy klasyfikacji. Zakłada się tam, że zmienne są dobrane w sposób poprawny, czyli że spełniają kryteria merytoryczne. Kryteria formalne są pomijane.” Patrz zarówno klasyczne prace na ten temat, takie jak np. Sokal i Sneath (1963), Anderberg, M. R. (1973), Hartigan (1975) jak i bardziej współczesne Everitt i in. (2001), Koronacki J. i Ćwik J. (2005), Krzyśko i in. (2008).

³² Gorzelak (1979, s. 20) zaliczył te własności do grupy kryteriów formalnych, przy czym przypomnijmy Gorzelak nie rozróżnia własności wg kryterium statystycznego jak Grabiński. Kryteria dzieli tylko na dwie grupy: merytoryczną i formalną.

nych w badaniach taksonomicznych, a więc szeroko rozumianych zagadnieniach klasyfikacji. Rozważania Grabińskiego dotyczą, przede wszystkim, badań z zakresu WAP³³ a więc konstrukcji miar syntetycznych. W tych badaniach na cechy narzuca się znacznie silniejsze ograniczenia i własności 10 i 12 u Grabińskiego związane są, przede wszystkim, z problematyką porządkowania liniowego. Własność dziesiąta w kontekście ogólnej klasyfikacji (nieliniowej) jest niezbyt zrozumiała, a dwunasta jest niepotrzebna.

Kryterium formalne nie zostało przez żadnego z przytaczanych autorów zdefiniowane i często jest łączone ze statystycznym³⁴. Zostaliśmy jednak przy jego wyodrębnieniu z powodów, o których będzie mowa poniżej.

Według tego kryterium, cechy powinny mieć własności: mierzalności (Fierich, Gorzelak, Grabiński, Kukuła (2000)), kompletności danych, ciągłości rozumianej przez Grabińskiego jako konieczność eliminacji zmiennych specyficznych uniemożliwiających wzajemną porównywalność obiektów, zaś wg Gorzelaka w sensie matematycznym³⁵.

Jak stwierdza Grabiński (1992, s.29 i nast.), spełnienie kryterium merytorycznego i formalnego jest obligatoryjne, aczkolwiek „Ich ocena dokonywana przez prowadzącego badania (ekspertów) jest subiektywna i trudno ją zalgorytmizować.”

Kryteria statystyczne „(...) opierają się na miernikach obiektywnych i są najczęściej podstawą konstruowania algorytmów eliminowania zmiennych o małym stopniu diagnostyczności. Kryteriów tych nie traktuje się zwykle jako obligatoryjne, a wynikający z nich zestaw zmiennych diagnostycznych jest zwykle kompromisem, między liczbą zmiennych, a liczbą kryteriów diagnostyczności.” (Grabiński 1992, s.29-30)

Do tych kryteriów Grabiński zalicza kryteria siódme do dziewiątego oraz „wraca” do postulatu piątego Fiericha, zaliczając do kryterium statystycznego. Bardzo wyraźnie widać, że Grabiński określając poszczególne typy kryteriów i „przydział” tychże do poszczególnych typów ma na myśli wąski typ badania – klasyfikację wg jakiejś miary syntetycznej.

³³ Wielowymiarowa analiza porównawcza.

³⁴ Wymieniony już Gorzelak (1979), także np. Frankowski (1991), z kolei Kukuła (2000) w ogóle nie wyodrębnia jakichkolwiek kryteriów. Grabiński w pracy z 1992 kryteria statystyczne traktuje jako podzbiór w zbiorze kryteriów formalnych, dopiero w pracy Grabiński (1992) traktuje je odrębnie.

³⁵ Gorzelak rozumie ciągłość w sensie matematycznym i postuluje by eliminować z badania cechy skokowe. Jest to postulat techniczny związany z techniką badania taksonomicznego. W tej chwili należy go uznać za nieaktualny w związku z rozwojem metod taksonomicznych, które można stosować do cech (zmiennych) mierzonych w skalach słabych (nominalnych i rangowych). Patrz np. monografie Gatnara (1998, 2003).

Podsumowując problem, własności, które powinien spełniać zbiór cech diagnostycznych można podzielić na trzy rodzaje wg kryteriów: merytorycznego, formalnego i statystycznego, w tym zakresie można uznać, że panuje zgoda autorów. Większym problemem jest zaliczenie poszczególnych własności do poszczególnych typów.

Biorąc pod uwagę badanie typu taksonomicznego w szerokim tego słowa znaczeniu cechy, można uznać, że obligatoryjnie cechy powinny mieć *własności merytoryczne*, do których zaliczymy własności F1-F3 (bez wymogu mierzalności, który zaliczymy do własności formalnych). Do obligatoryjnych można także zaliczyć *część własności formalnych*: mierzalność cech, ich duży zakres zmienności³⁶, nieco dyskusyjny jest wymóg reprezentatywności cech względem cech odrzuconych, czy jest to wymóg formalny czy merytoryczny. Pozostałe własności można określić jako *formalno-statystyczne* i ich spełnienie jest pożądane lub - przy badaniach opartych o metody analizy czynnikowej lub miary syntetyczne - wręcz wymagane. Są to: brak skorelowania (lub słabe) cech między sobą, ale skorelowanie ze zmiennymi nieuwzględnionymi; wysokie skorelowanie z czynnikami głównymi, wyodrębnionymi w całym zbiorze analizowanych zmiennych; dodatni lub negatywny związek z konstruowaną zmienną wyjścia³⁷, liniowość związków pomiędzy zmiennymi.

Piśmiennictwo

1. Anderberg, M. R., 1973. *Cluster analysis for applications*. New York: Academic Press.
2. Andrews, Harry C., 1972. *Introduction to mathematical techniques in pattern recognition*, Wiley J., New York.
3. Balicki A., 2009. *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
4. Borys T., 1984. *Kategoria jakości w statystycznej analizie porównawczej*. Prace Naukowe AE we Wrocławiu 284, seria: Monografie i Opracowania 23, Wrocław, s. 182-198.
5. Chojnicki Z., Czyż T., 2006. *Aspekty regionalne gospodarki opartej na wiedzy w Polsce*. Bogucki Wydawnictwo Naukowe, Poznań.
6. Everitt, B. S., Landau, S., Leese, M., 2001. *Cluster analysis* (4th Ed.). Arnold, London.

³⁶ Jest to własność określająca zdolność cechy do dyskryminacji zbioru obiektów, jest to nieco ogólniej sformułowany postulat F4. Fierich w swej pracy prezentował badanie oparte na jednostkach przestrzennych (badań rejonizację).

³⁷ Chodzi o konieczność określenia czy cechy są stymulantami czy destymulantami w badaniach porównawczych opartych o miary syntetyczne.

7. Fajferek A., 1966. *Region ekonomiczny i metody analizy regionalnej*. PWE, Warszawa.
8. Fierich J., 1957. *Próba zastosowania metod taksonomicznych do rejonizacji systemów rolniczych w województwie krakowskim*. Myśl Gospodarcza 1.
9. Frankowski Z., 1991. *Zastosowanie metod taksonomicznych w badaniach przestrzennych*, Agencja Wydawnicza Instytutu Gospodarki Przestrzennej i Komunalnej, Warszawa.
10. Gatnar E. (red.), 2003. *Analiza i prognozowanie zjawisk rynkowych o charakterze niometrycznym*, Wydaw. Uczelniane Akademii Ekonomicznej w Katowicach, Katowice.
11. Gatnar E., 1998. *Symboliczne metody klasyfikacji danych*, Wydaw. Naukowe PWN, Warszawa.
12. Gorzelak G., 1979. *Dobór zmiennych w statystycznej analizie porównawczej – metody (I)*. Wiadomości Statystyczne 3, s. 17-21.
13. Grabiński T., 1988. *Metody statystycznej analizy porównawczej*. (w:) A. Zeliaś (red.). *Metody statystyki międzynarodowej*. PWE, s. 235-259.
14. Grabiński T., 1991. *Metody aksometrii*. Akademia Ekonomiczna w Krakowie.
15. Grabiński T., 1992. *Podstawy kwantyfikacji zmiennych przestrzennych*. (w:) S. Mynarski (red.). *Badania przestrzenne rynku i konsumpcji*. Wydawnictwo Naukowe PWN, s. 25-35.
16. Grabiński T., Wydymus S., Zeliaś A., 1989. *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*. PWN, Warszawa.
17. Hartigan J.A., 1975. *Clustering Algorithms*. John Wiley and Sons, New York.
18. Hellwig Z., 1968. *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*. Przegląd Statystyczny 4.
19. Jajuga K., 1990. *Statystyczna teoria rozpoznawania obrazów*. Wydaw. Naukowe PWN, Warszawa.
20. Jajuga K., 1993. *Statystyczna analiza wielowymiarowa*. Wydaw. Naukowe PWN, Warszawa.
21. Koronański J., Ćwik J., 2005. *Statystyczne systemy uczące się*. Wydawnictwa Naukowo-Techniczne, Warszawa.
22. Kosiedowski W., Kufel T., Popławski W., 1989. *Regionalizacja województwa wrocławskiego za pomocą metod taksonomicznych*. Wiadomości Statystyczne 3.
23. Kowerski M., 1983. *Kilka uwag na temat doboru zmiennych charakteryzujących poziom rozwoju gospodarczego województw*. Wiadomości Statystyczne 11, s.33-34.
24. Krzyśko, M., Wołyński W., i in., 2008. *Systemy uczące się: rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości*. WN-T, Warszawa.
25. Kukuła K., 2000. *Metoda unitaryzacji zerowanej*. Wydawnictwo Naukowe PWN, Warszawa.
26. Nowak E., 1990. *Metody taksonomiczne w klasyfikacji obiektów gospodarczych*. PWE, Warszawa.

27. Panek T., 2009., *Statystyczne metody wielowymiarowej analizy porównawczej*. Oficyna Wydawnicza SGH w Warszawie, Warszawa.
28. Parysek J. J., 1982. *Modele klasyfikacji w geografii*. Uniwersytet im. A. Mickiewicza w Poznaniu, Poznań.
29. Pluta W., 1972. *Grafowa metoda klasyfikacji cech*. Prace Naukowe WSE 21-22, Wrocław.
30. Pocięcha J., 2008. *Rozwój metod taksonomicznych i ich zastosowań w badaniach społeczno-ekonomicznych*. Publikacja z konferencji naukowej z okazji obchodów jubileuszu 90-lecia GUS w 2008 roku „Statystyka Społeczna. Dokonania-Szanse-Perspektywy”. Dokument elektroniczny. Tryb dostępu: www.stat.gov.pl/gus/5840_4193_PLK_HTML.htm, stan z dn. 13.10.2010 r.
31. Pocięcha J., Podolec B., Sokołowski A., Zając K., 1988. *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, Warszawa, Państwowe Wydawnictwo Naukowe.
32. Podolec B., Zając K., 1978. *Ekonometryczne metody ustalania rejonów konsumpcji*. PWE, Warszawa.
33. Polak H., Starzewska Ł., 1975. *System wskaźników rozwoju społeczno-gospodarczego do porównań jednostek terytorialnych*. Seria Z prac ZBSE 95, GUS, Warszawa.
34. Romesburg, H. C., 2004. *Cluster analysis for researchers*. Lulu, North Carolina.
35. Rozin B. B., 1979. *Teoria rozpoznawania obrazów w badaniach ekonomicznych*. PWN, Warszawa.
36. Rutkowski T., 1982. *Teoria i praktyka badań empirycznych. (Uwagi na marginesie artykułu Aleksandry Kani-Gospodarowicz)*. Wiadomości Statystyczne 6, s.26-27.
37. Sneath P.H.A., Sokal R.R., 1963. *The Principles of Numerical Taxonomy*. W. H. Freeman Ltd., San Francisco.
38. Tymowska K., 1979. *W sprawie mierzenia poziomu rozwoju gmin – uwagi krytyczne*. Wiadomości Statystyczne 9, s.22-24.