316/2007

Raport Badawczy
Research Report

RB/74/2007

Integracja metod oceny nauczycieli i studentów dla potrzeb konstrukcji modelu szkoły wyższej

M. Bereziński, M.Inkielman, D.Wagner

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
Prof. dr inż. Roman Kulikowski

Warszawa 2007

INTEGRACJA METOD OCENY NAUCZYCIELI I STUDENTÓW DLA POTRZEB KONSTRUKCJI MODELU SZKOŁY WYŻSZEJ

Mirosław Bereziński[1], Michał Inkielman[2], Dariusz Wagner[3]
Instytut Badań Systemowych PAN
01-447 Warszawa, ul. Newelska 6
e-mail: [1]Miroslaw.Berezinski@ibspan.waw.pl
[2]Michal.Inkielman@ibspan.waw.pl
[3]Dariusz.Wagner@ibspan.waw.pl

Streszczenie. Praca dotyczy zastosowań metod statystycznych do oceny jakości procesu edukacyjnego. Tradycyjne metody oceniania traktują rozdzielnie proces nauczania i uczenia się. W pracy zwrócono uwagę na niesystemowość tego podejścia i przedstawiono koncepcję oceniania zintegrowanego, w którym uczestniczą obie strony procesu edukacyjnego: studenci i nauczyciele. U podstaw koncepcji leży sformułowany przez autorów postulat wieloaspektowej systemowej jednorodności grup zajęciowych oraz racjonalności kojarzenia odpowiednich nauczycieli akademickich z odpowiednimi grupami zajęciowymi. Przyjęto, że wśród cech charakteryzujących zbiorowości studentów bądź nauczycieli istnieje cecha priorytetowa z punktu widzenia celu badania. Sformułowano kryterium jednorodności. Ma ono postać wielowymiarowej statystyki, której rozkład jest tym bliższy rozkładowi chi-kwadrat, im liczniejsza jest rozpatrywana zbiorowość. Przedstawiono i przedyskutowano metodę dekompozycji niejednorodnej zbiorowości studentów (nauczycieli) na grupy jednorodne. Przeanalizowano rzeczywiste uwarunkowania procesu oceniania studentów oraz nauczycieli akademickich. Zwrócono uwagę na brak równowagi tego procesu i przedstawiono propozycję metody identyfikacji przyczyn i skutków tego zjawiska. Wskazano na możliwość poprawy jakości procesu edukacyjnego przez racjonalizację przyporządkowywania nauczycieli grupom zajęciowym. Sformułowano wstępną koncepcję systemowego kryterium jakości funkcjonowania szkoły wyższej, opartą na wykorzystaniu odpowiednio sformułowanej funkcji użyteczności, jako miary strat i ryzyka wynikających z niewłaściwego zorganizowania procesu edukacyjnego.

Słowa kluczowe: proces edukacyjny, szkoła wyższa, ocena nauczycieli, ocena studentów, statystyka matematyczna, niejednorodność, nierównowaga, użyteczność

1. Wprowadzenie

Badania nad modelowaniem procesu edukacyjnego realizowanego w szkole wyższej są przez autorów prowadzone od kilku lat. Dotychczasowe wyniki badań są opublikowane (Bereziński 2003; Bereziński, Inkielman i Wagner 2004, 2005a, 2005b, 2006). Koncentrowały się one analizie rzeczywistej struktury, organizacji i sposobu realizacji procesu edukacyjnego w szkole wyższej oraz na skonstruowaniu takiego matematycznego modelu tego procesu, który uwzględniałby jego dynamikę, niepewność i nieokreśloność. Model miał postać hierarchicznej sieci stochastycznej, zaś proces przechodzenia studentów przez kolejne fazy procesu edukacyjnego był opisany za pomocą niejednorodnego łańcucha Markowa (Bereziński 2003; Bereziński, Inkielman i Wagner 2004, 2005b). Istotnym aspektem badań było zwrócenie uwagi na konieczność traktowania procesu edukacyjnego jako systemowej jedności, której dwoma podstawowymi elementami są studenci oraz nauczyciele akademiccy. Ważne też było podkreślenie, że chociaż rola każdego z tych elementów w procesie edukacyjnym jest różna (student jest osobą uczącą się, nauczyciel – osobą nauczającą), to jednak cel główny jest wspólny: jest nim wypromowanie absolwenta mającego taki zasób wiedzy specjalistycznej i taką umiejętność jej stosowania, które w pełni odpowiadają jego możliwościom oraz aspiracjom i czynią zeń fachowca poszukiwanego na konkurencyjnych rynkach pracy. Zwrócono również uwagę na konieczność prowadzenia procesu edukacyjnego w sposób gospodarny, a więc racjonalny ekonomicznie i technicznie (Bereziński, Inkielman i Wagner 2005a). Stwierdzono, że proces ten powinien mieć odpowiednią jakość: jego głównym celem jest bowiem wykształcenie dobrze wykwalifikowanych specjalistów, którzy – w warunkach konkurencji panującej na krajowych i zagranicznych rynkach pracy – bez trudu znajdą zatrudnienie zgodne z ukończonym kierunkiem studiów i odpowiadające ich predyspozycjom oraz aspiracjom. Wyrażono przekonanie, że aby to było możliwe, potrzebne jest stałe monitorowanie procesu edukacyjnego i ustawiczne ocenianie jego jakości. Ponieważ o jakości tej decydują przede wszystkim studenci i nauczyciele, więc zaistniała konieczność opracowania metod oceny obu tych podmiotów. Metody takie zostały przedstawione i przedyskutowane w raporcie ubiegłorocznym (Bereziński, Inkielman i Wagner 2006).

Zadanie badawcze, na którym pracowano w 2007 r., dotyczyło integracji metod oceny nauczycieli i studentów. U źródeł zadania leży fakt, że struktura tradycyjnie stosowanego w edukacji procesu oceniania jest sprzeczna z zasadą systemowości procesu edukacyjnego: w procesie tym nauczyciele oceniają studentów, sami są oceniani przez władze szkoły, zaś szkoła jest oceniana przez odpowiednie czynniki zewnętrzne, np. przez Państwową Komisję Akredytacyjną. Ta jednokierunkowa hierarchiczna struktura odpowiada przedmiotowemu traktowaniu studenta (jako obiektu, który jest nauczany i oceniany), nauczyciela (jako obiektu, który uczy i też podlega ocenie) oraz szkoły. Jest to podejście niewłaściwe. Na każdym etapie procesu edukacyjnego występują bowiem dwa podmioty (studenci oraz nauczyciele akademiccy, nauczyciele oraz władze szkoły, szkoła oraz nadrzędna jednostka oświatowa), które wzajemnie na siebie oddziałują i wzajemnie się kształtują. W pracy tej będziemy się zajmować tylko pierwszym z tych etapów. Wychodzimy z przekonania, że chociaż statutowym obowiązkiem nauczycieli jest ocenianie stanu wiedzy studentów, to jednak niezbywalnym prawem studentów jest wyrażanie opinii o tych. którzy ich uczą, egzaminują i wychowują. Innymi słowy, procesy oceniania studentów i nauczycieli nie powinny biec rozłącznie, lecz stanowić jeden spójny, zintegrowany system.

Praca jest więc poświęcona analizie i – częściowo – syntezie procesu oceniania, jako istotnego czynnika kształtowania systemowej jakości procesu edukacyjnego w szkole wyższej. Jakość ta zależy w dużej mierze od jakości relacji zachodzących między studentami i nauczycielami, od psychofizycznych cech jednych i drugich, od stosunku studentów do nauki, od merytorycznego i pedagogicznego przygotowania nauczycieli itd. Relacje te nie są do końca znane. Przyjęto więc, że są obciążone niepewnością i jako model niepewności przyjęto przypadkowość probabilistyczno-statystyczną.

Praca jest zorganizowana w następujący sposób. W rozdziale 2 przedstawiono ogólny kontekst zagadnienia, zwracając szczególną uwagę na systemową jedność procesu edukacyjnego, z której – między innymi – wynika systemowa jedność procesu oceny studentów i nauczycieli. W rozdziale 3 przyjęto, że systemowa struktura procesu oceny jakości edukacji może być odzworowana za pomocą odpowiedniej macierzy – przedstawiono tę macierz i przeprowadzono jej dyskusję. Rozdział 4 poświęcono zastosowaniu statystycznych metod badania jednorodności wielowymiarowych zbiorów danych do analizy zbiorów danych zakresu oceny procesu edukacyjnego. W szczególności omówiono metodę Abbego-Helmerta i podstawowe własności związanej z nią statystyki Q. Wskazano zalety oraz wady tej metody. Zwrócono uwagę na to, że podział zbioru danych na podzbiory jednorodne może być uważany za jedną z realizacji algebraicznej zasady identyfikacji elementów równoważnych. W żwiązku z tym, uwzględniając specyfikę rozpatrywanego w tej pracy problemu, wprowadzono zgodne z nią pojęcie równoważności zmiennych losowych, a następnie wykorzystano je do sformułowania warunków statystycznej jednorodności zbiorów jednowymiarowych i wielowymiarowych oraz warunków statystycznej jednorodności zbiorowości składającej się z wielu zbiorów, wobec których istnieje niepewność co do ich jednorodności. Przedstawiono procedurę dekompozycji wielowymiarowego zbioru danych na grupy jednorodne w sensie wektorowego wskaźnika jakości. W rozdziale 5 podano przykłady ilustrujące sposób wykorzystania przedstawionych metod do testowania jednorodności zbiorowości studentów bądź nauczycieli akademickich, ocenianych za pomocą kryteriów wielowymiarowych. Rozdział 6 poświęcono omówieniu relacji między klasyczną statystyką T^2 Hotellinga badania jednorodności a statystyką U. W rozdziale 7 zwrócono uwagę na możliwość racjonalizacji procesu edukacyjnego przez optymalizację przydziału nauczycieli akademickich do studenckich grup zajęciowych. W szczególności podano wstępną koncepcję użycia funkcji użyteczności do sformułowania kryterium efektywności rozwiązania tego zadania. W rozdziale 8 zwrócono uwagę na systemową nierównowagę procesu edukacyjnego. Zaproponowano wstępną koncepcję metody identyfikacji przyczyn i skutków braku równowagi. Omówiono też czynniki, za pomocą których można sprowadzać proces edukacyjny do stanu równowagi. Rozdział 9 zawiera sformułowanie uwag końcowych i wniosków.

2. Ogólny kontekst zagadnienia

Nieustannie podkreśla się, że na ogólny obraz stanu osobowego i społecznego rozwoju człowieka składa się wiele czynników, między innymi poziom i jakość jego edukacji. Zależą one, przede wszystkim, od poziomu i jakości całego łańcucha dydaktyczno-wychowawczego, w którym uczestniczy człowiek (wychowanie przedszkolne, edukacja wczesnoszkolna, szkoła podstawowa, gimnazjum, liceum ogólnokształcące lub zawodowe, studia wyższe, studia podyplomowe, studia doktoranckie itd.). Od szeregu lat władze edukacyjne i parlamenty wielu krajów, które w przeszłości szczyciły się niezwykle wysokim poziomem rozwoju nauki i techniki, władze szkół wyższych, rodzice i coraz częściej sami uczniowie i studenci niezmiernie krytycznie oceniają jakość edukacji (Hanoushek 1973; Perkins 1974; Ravitch 1981; Tate 1993; Elmore 1995; Hanoushek 1996; Reid 1999; McAlpine i Weston 2000; Girod 2002). Powszechnie dostrzega się spadek poziomu nauczania na wszystkich szczeblach szkolnictwa, obniżanie się stopnia osiągnięć, inteligencji, erudycji, wartości, umiejętności. Chociaż głosi się hasła gloryfikujące kapitał intelektualny, to jednak poziom edukacji i społeczne kryteria oceny jej jakości są tak silnie zdeformowane, że wielu krajom grozi zapaść cywilizacyjna. Kraje bogate ratują się importem specjalistów z krajów biedniejszych, które wskutek tego tracą swoje elity umysłowe i obniżają własny potencjał rozwojowy.

Przyczyn tego stanu rzeczy jest wiele, ale nie ma najmniejszych wątpliwości, że jedną z podstawowych jest zdeprecjonowanie w wielu krajach roli nauk ścisłych, a zwłaszcza matematyki, jako ważnego elementu ogólnej kultury osobistej i społecznej człowieka. Przykładem może być Francja, która na przestrzeni wieków dała Europie i światu rzesze wybitnych matematyków i inżynierów, a obecnie poziom matematycznej i technicznej edukacji społeczeństwa francuskiego jest tak niski, że – wedle oficjalnych oświadczeń władz szkolnych tego kraju – musi być uznany za katastrofalny (Dorier 2001). Podobna sytuacja ma miejsce w innych krajach (ICME 2004). Nie inaczej jest w Polsce – poziom wiedzy matematycznej przeciętnego wykształconego członka społeczeństwa jest bardzo niski, a rektorzy szkół wyższych biją na alarm, ponieważ młodzież zgłaszająca się na studia wyższe ma tak słabe przygotowanie matematyczne, że nie sposób realizować obowiązujących programów studiów. W wielu krajach zwraca się uwagę, że – mimo wszystkich zalet Internetu – niewłaściwe korzystanie z tego systemu przez młodzież szkolną i akademicką jest jednym z głównych czynników obniżenia się poziomu edukacji.

Obecnie w niemal wszystkich krajach podejmuje się próby uzdrowienia szkolnictwa i to nie tylko w sferze zakresu i metod nauczania matematyki. Kraje Unii Europejskiej nie są tu wyjątkiem. Powszechnie podkreśla się, że jedną z głównych przyczyn kryzysu edukacyjnego jest upadek autorytetu nauczyciela, jako społecznie uznanej powagi w dziedzinie edukacji i wychowania. Trudno się z tym nie zgodzić. Trzeba jednak zdawać sobie sprawę, że upadek ten został spowodowany przez niewłaściwe doktryny pedagogiczne, które w większości krajów opanowały szkolnictwo wszystkich szczebli zwłaszcza w drugiej połowie 20. wieku, a które zostały oparte na fałszywym postulacie, że wszelki autorytet jest szkodliwy w edukacji i wychowaniu, niszczy bowiem wolność jednostki i kształtuje wśród młodzieży postawy uległości, podporządkowania oraz niesamodzielności. Nauczyciel przestał więc być dla ucznia i studenta autorytetem – pojęcie autorytetu zastąpiono pojęciami przyjaźni, koleżeństwa, współpracy, partnerstwa itp. A przecież pojęcie autorytetu w nauce i edukacji dotyczy relacji międzyosobowych i wiąże się z hierarchicznym oraz wieloaspektowym sposobem istnienia oraz organizacji społeczeństwa(Bocheński 1993). Nauczyciel, który miał opinię autorytetu, nie mógł nie być przyjacielem dzieci i młodzieży, musiał w pracy edukacyjnej respektować właściwie rozumiane zasady współpracy i partnerstwa, ale musiał też umieć właściwie ustawić hierarchię relacji między sobą, a tymi których uczył i wychowywał. Na autorytet nauczycielski trzeba było bowiem zapracować i aby go nie stracić należało ustawicznie doskonalić swoją osobowość i charakter, trzeba było stale pogłębiać, poszerzać i uaktualniać swoją wiedzę przedmiotową i ogólną. Temu właśnie służyły rozmaite formy kształcenia i dokształcania nauczycieli (studia podstawowe, uzupełniające i podyplomowe, konferencje naukowe, seminaria, spotkania wszechnicowe, kursy itp.). Nauczyciel, którego pozbawiono funkcji autorytetu, musiał znacznie zmienić i inaczej ukierunkować wymagania w stosunku do samego siebie i swoich wychowanków, z oczywistą szkodą dla jakości edukacji. Podstawowym warunkiem poprawy jakości edukacji jest więc odbudowa właściwie rozumianego autorytetu nauczyciela jako dydaktyka i wychowawcy.

Mimo wypaczenia roli i funkcji nauczyciela słusznie uważa się, że jakość edukacji zależy przede wszystkim od nauczycieli. Jest to fakt oczywisty: cała historia szkolnictwa uczy, że nazwisko, autorytet i sława nauczyciela przyciąga do uczelni młodzież, która chce się kształcić pod jego kierunkiem. W masie szkół wyższych, które powstały w ostatnich latach, niewielu jest takich nauczycieli, ale rozwijają się i zyskują rozgłos przede wszystkim te z nich oraz te dyscypliny naukowe i techniczne, które kojarzą się z nazwiskami uznanych autorytetów, twórców szkół naukowych i dydaktycznych. Wiadomo, na przykład, jaką sławę zdobyły uniwersytety związane z nazwiskami R. Feynmana (University of Princeton), I. Prigogine'a (Université Libre de Bruxelles oraz University of Texas) oraz B. Mandelbrota (Université de Lille) po otrzymaniu przez tych naukowców nagrody Nobla i jak silnie fakt ten wpłynął na zainteresowanie młodzieży studiami w tych uczelniach, zwłaszcza w dziedzinie fizyki. Powstały słynne szkoły naukowe Feynmana, Prigogine'a i Mandelbrota, a ich absolwenci byli przez kilka dziesięcioleci najlepszą wizytówką i reklamą swoich uniwersytetów.

Powstaje jednak wątpliwość, czy rzeczywiście – a jeśli tak, to w jakiej mierze – wyniki uzyskiwane przez absolwentów szkół są odbiciem jakości ich nauczycieli. Jak bowiem należałoby oceniać wielu nauczycieli zatrudnionych w szkolnictwie niższego stopnia, skoro tak olbrzymią popularnością cieszą się korepetycje oraz kursy przygotowujące młodzież gimnazjalną do egzaminu końcowego, a młodzież licealną do matury? Powszechność tego zjawiska podważa tezę o dobrej jakości systemów edukacyjnych w wielu krajach (zob., np., Soar, Medley i Coker 1983; Tang 1988; Kupisiewicz 1984). Źle przygotowani absolwenci gimnazjów w większości wybierają w liceach te profile edukacyjne, które wymagają uczenia się matematyki na poziomie podstawowym, a źle przygotowani absolwenci liceów mają kłopoty na tych kierunkach studiów, które wymagają posługiwania się matematyką (studia techniczne, matematyka, fizyka, informatyka, ekonomia, zarządzanie, marketing itd.).

Dopóki nie zostanie uzdrowiona sytuacja w szkolnictwie niższym, szkoły wyższe są zmuszone do takiej reorganizacji prowadzonego w nich procesu edukacyjnego, która stworzy słabo przygotowanej młodzieży możliwość uzupełnienia braków w stopniu umożliwiającym odbywanie studiów. W związku z tym powstało zapotrzebowanie na nowy typ nauczyciela akademickiego, posiadającego doświadczenie z pracy zarówno w szkolnictwie niższym, jak i wyższym, znającego programy przedmiotowe oraz metody dydaktyczno-pedagogiczne obowiązujące w szkolnictwie niższym i potrafiącego wykorzystać je do przeprowadzenia zajęć wyrównawczych bądź uzupełniających z młodzieżą rozpoczynającą studia. Jest to olbrzymi wysiłek ze strony szkół wyższych, ale opłacalny z punktu widzenia wykonywania przez nie ich misji oraz z punktu widzenia całej kultury i edukacji narodowej. Wysiłek ten musi być wpleciony w proces zarządzania szkołą i kształtowania jej rozwoju w różnych horyzontach czasowych. Jest to złożony, wieloaspektowy i wielokontekstowy proces decyzyjny, realizowany w warunkach niepełności, niepewności i niedokładności danych oraz nieokreśloności przyszłości. Czynniki te wyznaczają obszar poszukiwań modelu szkoły, który będzie pomocny w podejmowaniu decyzji operacyjnych, taktycznych i strategicznych. Głównym systemowym elementem modelu szkoły wyższej musi być model prowadzonego w niej procesu edukacyjnego. Zwracamy jeszcze raz uwagę na to, że w procesie tym występują i wzajemnie na siebie oddziałują dwie zbiorowości ludzkie – studenci oraz nauczyciele akademiccy. Aby proces ten odbywał się zgodnie z obowiązującymi standardami programowymi i dawał zamierzone wyniki, konieczna jest jego ciągła kontrola i zwrotne wykorzystywanie zdobytych w ten sposób wiedzy i doświadczenia do kształtowania jego jakości.

Jakość procesu edukacyjnego zależy od wielu czynników materialnych i niematerialnych, ale podstawowe znaczenie mają te z nich, których nośnikami są zbiorowości uczniów i nauczycieli. W tradycyjnym schemacie edukacji nauczyciel jest nie tylko wychowawcą i stroną przekazującą studentom wiedzę oraz inspirującą do jej zdobywania, pogłębiania i rozszerzania, ale także osobą oceniającą poziom i jakość wiedzy przedmiotowej studentów oraz umiejętność jej stosowania. Z kolei władze uczelni oceniają nauczyciela, biorąc pod uwagę przede wszystkim jego status naukowy, liczbę wypromowanych absolwentów, liczbę prowadzonych prac przejściowych, końcowych i dyplomowych, liczbę publikacji, uczestnictwo w konferencjach naukowych itp. W ocenach tych na ogół nie bierze się pod uwagę opinii, jaką o swoich nauczycielach mają studenci. Tymczasem o jakości nauczyciela akademickiego nie świadczą tylko jego stopnie i tytuły naukowe, ani też jego dorobek publikacyjny itp., lecz przede wszystkim efekty prowadzonych przez niego zajęć dydaktycznych, a więc wyniki osiągane przez studentów. Nauczyciel akademicki może posiadać wielki dorobek naukowy, ale nie mieć umiejętności przekazywania wiedzy, nie potrafić przykuć uwagi słuchaczy. Może egzekwować wiedzę w sposób zbyt powierzchowny lub zbyt głęboki itd. Szczególnie widoczne jest to w przypadku wielu nauczycieli matematyki, nie mających jednak wykształcenia matematycznego. Brak zrozumienia ducha matematyki sprawia, że często przykładają oni większe znaczenie do pamięciowego opanowania przez studenta określonych partii materiału niż do wyrobienia umiejętności myślenia i wnioskowania w kategoriach matematycznych. Student może umieć poprawnie sformułować i udowodnić określone twierdzenie matematyczne, nie rozumiejąc ani jego istoty, ani potrzeby jego znajomości, ani też nie widząc obszarów jego możliwych zastosowań. Ważne jest więc, aby nauczyciel akademicki miał nie tylko odpowiedni zasób wiedzy przedmiotowej, ale by posiadał umiejętności dydaktyczne i właściwe przygotowanie pedagogiczne.

Nie każdy dobry naukowiec jest dobrym dydaktykiem. Sytuację tę pogłębia fakt, że wielu nauczycieli akademickich nie ma przygotowania pedagogicznego. Nauczyciel bez tego przygotowania jest bardziej narażony na popełnianie istotnych błędów pedagogicznych niż osoba właściwie przygotowana do pracy nauczycielskiej. Konflikty między nauczycielami i studentami, w których z reguły przegrywa student – a opinia ta szybko rozchodzi się wśród młodzieży ubiegającej się o przyjęcie na studia wyższe – są głównym powodem tego, że szkoły o nawet bardzo wysokim poziomie merytorycznym tracą z roku na rok studentów, którzy wolą wybrać szkołę mniej renomowaną, ale rzeczywiście im przyjazną. Obserwowany obecnie kryzys wielu szkół wyższych lub wydziałów jest w większości przypadków spowodowany tym, że kadra dydaktyczna nie ma przygotowania pedagogicznego. Nie do rzadkości należą przypadki, gdy student, zaliczywszy pierwszy lub drugi rok studiów, zmienia szkołę lub kierunek edukacji z powodu niemożności zaliczenia ćwiczeń lub zdania egzaminu z jakiegoś przedmiotu. Są to fakty na tyle znane i doświadczane w szkolnictwie wyższym, że nie wymagają szerszego objaśniania. Skłaniają jednak do wyrażenia poglądu, że nie jest właściwe ograniczenie się w procesie edukacyjnym tylko do oceny studentów przez nauczycieli. Studenci powinni mieć prawo oceniania swych nauczycieli, a ich opinie powinny być brane pod uwagę przez władze szkoły wyższej. Oczywiście, proces oceniania nauczycieli przez studentów nie może odbywać się bez wiedzy nauczycieli. Niedopuszczalne i godne pożałowania są wciąż jeszcze spotykane sytuacje, kiedy to opinie o prowadzącym zajęcia są zbierane bez jego wiedzy. Pomijając niewychowawczy aspekt takiego postępowania trzeba stwierdzić, że szkoła, w której tak się czyni, sama niszczy swoją renomę przez podważanie autorytetu nauczyciela akademickiego.

Oceny studentów dokonuje się w oparciu o tzw. pomiar sprawdzający zwany też pomiarem dydaktycznym (Niemierko 1975, 1990, 1991). W języku teorii identyfikacji jest to odpowiednik świadomie i celowo zaprojektowanego i przeprowadzonego eksperymentu identyfikacyjnego. Najbardziej rozpowszechnioną metodą pomiaru sprawdzającego jest kolokwium albo egzamin pisemny (tradycyjny lub w formie testu) bądź ustny, w wyniku których otrzymuje się ilościowe oszacowania wielkości i jakości wiedzy, umiejętności, predyspozycji, cech i innych charakterystyk posiadanych lub nabytych przez studentów. Podstawowe wyniki pomiaru są dokumentowane w formie kart egzaminacyjnych oraz protokołów i archiwizowane.

W odniesieniu do kadry nauczycielskiej dane o rozwoju zawodowym nauczyciela też są systematycznie gromadzone i wykorzystywane dla poprawy jakości procesu edukacyjnego. Dane te dostarczają władzom szkoły użytecznej informacji o jakości pracy nauczycieli, stanowią wyraz istnienia sprzężenia zwrotnego między władzami szkoły a nauczycielem (w oparciu o nie przekazuje się nauczycielowi zalecenia lub instrukcje dotyczące sposobów poprawy jakości nauczania, urealnia się cele edukacji, ocenia się stopień spełnienia celów, stwarza warunki do poprawy umiejętności zawodowych nauczyciela itp.), a także pozwalają na ciągłe monitorowanie stanu procesu edukacyjnego (dzięki temu władze szkoły mogą podejmować działania sprzyjające utrzymaniu zawodowej jakości nauczycieli na wymaganym poziomie).

Chociaż ocenianie dotyczy zarówno studentów, jak i nauczycieli, to jednak ma charakter jednokierunkowy: władze szkoły oceniają nauczyciela, zaś nauczyciel ocenia studentów. Tymczasem, chociaż studenci mają swoje zdanie o poszczególnych nauczycielach, o sposobie prowadzenia przez nich zajęć, o sposobie przeprowadzania egzaminu, o metodach oceniania itp., to jednak opinie te nie są przez władze szkoły brane pod uwagę przy ocenach nauczycieli. Brak tego sprzężenia zwrotnego sprawia, że student jest wciąż traktowany bardziej w sposób przedmiotowy niż podmiotowy.

Odrębność oceniania jakości nauczycieli akademickich i studentów jest poważną niepoprawnością metodologiczną. Proces edukacyjny prowadzony w szkole wyższej jest bowiem celowo ukierunkowanym procesem systemowym i ocena któregokolwiek z jego elementów nie może być prowadzona w oderwaniu od oceny pozostałych. Mówiąc inaczej, ocena jakości nauczycieli akademickich nie może być prowadzona w oderwaniu od oceny jakości studentów i odwrotnie, ocena jakości studentów nie może być robiona niezależnie od oceny nauczycieli. Czasem próbuje się spełnić ten wymóg przyjmując, że jedną ze składowych procesu oceny nauczyciela jest opinia o nim wyrażona przez studentów, których uczy. Podejście to ma zarówno zwolenników, jak i przeciwników. Korzenie tego sporu mają przede wszystkim podłoże etyczne. Jedni uważają, że opinie wyrażane wzajemnie o sobie przez nauczycieli i studentów są obiektywne, bez względu na to, czy są pozytywne, czy negatywne. Drudzy twierdzą, że nauczyciel zawsze przestrzega zasad etyki zawodowej, zawsze kieruje się dobrem studenta i zawsze stara się oceniać go obiektywnie. Doświadczenie uczy, że oba te poglądy są zbyt radykalne: nierzadko jest tak, że nauczyciel przyjmuje każdą negatywną uwagę ze strony studenta jako atak na siebie lub że student w sposób tendencyjny negatywnie wypowiada się o nauczycielu. Fakty te nie podważają jednak zasadności poglądu, że skoro proces edukacyjny ma charakter systemowy, to w jego całościowej ocenie muszą uczestniczyć obie strony – nauczyciele i studenci – przy czym przy ocenianiu którejkolwiek z tych stron musi być brane pod uwagę zdanie strony przeciwnej.

Niech S będzie symbolem oceny dokonywanej przez studenta, N – symbolem oceny dokonywanej przez nauczyciela. Utwórzmy z tych symboli parę uporządkowaną o poprzedniku N i następniku S. Oznaczmy ją przez (N, S). Ponieważ zarówno ocena podana przez nauczyciela, jak i ocena podana przez studenta może być obiektywna lub nieobiektywna, więc są możliwe cztery następujące sytuacje:

- (N_+, S_+) – ocena wystawiona studentowi przez nauczyciela oraz ocena wystawiona nauczycielowi przez studenta są obiektywne,
- (N_+, S_-) – ocena wystawiona przez nauczyciela studentowi jest obiektywna, ale ocena wystawiona przez studenta nauczycielowi jest nieobiektywna,
- (N_-, S_+) – ocena wystawiona przez nauczyciela studentowi jest nieobiektywna, a ocena wystawiona przez studenta nauczycielowi jest obiektywna,
- (N_-, S_-) – obie oceny są nieobiektywne,

gdzie indeks „+" wskazuje na obiektywność, zaś indeks „-" na nieobiektywność oceny.

Oczywiście, najbardziej pożądana jest sytuacja (N_+, S_+), świadcząca o pełnej obiektywności ocen formułowanych przez studenta i nauczyciela, a najmniej pożądana sytuacja (N_-, S_-). Doświadczenie uczy, że w praktyce edukacyjnej zdarzają się wszystkie cztery sytuacje. Z tego powodu każda z nich powinna być uwzględniona w systemie oceny jakości procesu edukacyjnego.

3. Ogólna struktura macierzy oceny jakości procesu edukacyjnego

Skoro odrębne traktowanie procesów uczenia się i nauczania jest sprzeczne z zasadą systemowości edukacji, to również jest z nią sprzeczne odrębne ocenianie jakości tych procesów. Jakość procesu edukacyjnego musi być oceniana całościowo, a więc równocześnie z punktu widzenia jakości wiedzy i umiejętności stosowania zdobywanych przez studentów oraz z punktu widzenia jakości nauczania. Niech n_1 będzie liczbą studentów, n_2 – liczbą nauczycieli. Jeżeli symbolem $x_{i,j}^{(1)}$ $(i, j = 1,2,\ldots,n_1)$ oznaczymy oceny przyporządkowane przez studentów studentom, symbolem $x_{i,j}^{(2)}$ $(i = 1,2,\ldots,n_1;\ j = 1,2,\ldots,n_2)$ oceny przyporządkowane przez studentów nauczycielom, symbolem $x_{i,j}^{(3)}$ $(i=1,2,\dots,n_1;\ j=1,2,\dots,n_2)$ oceny przyporządkowane przez nauczycieli studentom, zaś symbolem $x_{i,j}^{(4)}$ $(i,j=1,2,\dots,n_2)$ oceny przyporządkowane nauczycielom przez nauczycieli, to zbiór wszystkich ocen można przedstawić w postaci następującej macierzy kwadratowej

$$\begin{pmatrix}
x_{1,1}^{(1)} & x_{1,2}^{(1)} & \dots & x_{1,n_1}^{(1)} & x_{1,1}^{(2)} & x_{1,2}^{(2)} & \dots & x_{1,n_2}^{(2)} \\
x_{2,1}^{(1)} & x_{2,2}^{(1)} & \dots & x_{2,n_1}^{(1)} & x_{2,1}^{(2)} & x_{2,2}^{(2)} & \dots & x_{2,n_2}^{(2)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
x_{n_1,1}^{(1)} & x_{n_1,2}^{(1)} & \dots & x_{n_1,n_1}^{(1)} & x_{n_1,1}^{(2)} & x_{n_1,2}^{(2)} & \dots & x_{s,n_2}^{(2)} \\
x_{1,1}^{(3)} & x_{1,2}^{(3)} & \dots & x_{1,n_1}^{(3)} & x_{1,1}^{(4)} & x_{1,2}^{(4)} & \dots & x_{1,n_2}^{(4)} \\
x_{2,1}^{(3)} & x_{2,2}^{(3)} & \dots & x_{2,n_1}^{(3)} & x_{2,1}^{(4)} & x_{2,2}^{(4)} & \dots & x_{2,n_2}^{(4)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
x_{n_2,1}^{(3)} & x_{n_2,2}^{(3)} & \dots & x_{n_2,n_1}^{(3)} & x_{n_2,1}^{(4)} & x_{n_2,2}^{(4)} & \dots & x_{n_2,n_2}^{(4)}
\end{pmatrix}. \qquad (3.1)$$

Jest to macierz blokowa o takiej strukturze

$$\begin{pmatrix}
X_{n_1\times n_1}^{(1)} & X_{n_1\times n_2}^{(2)} \\
X_{n_2\times n_1}^{(3)} & X_{n_2\times n_2}^{(4)}
\end{pmatrix}, \qquad (3.2)$$

gdzie $X_{n_1\times n_1 s}^{(1)}$ - macierz ocen studentów przez studentów, $X_{n_1\times n_2}^{(2)}$ - macierz ocen nauczycieli przez studentów, $X_{n_2\times n_1}^{(3)}$ - macierz ocen studentów przez nauczycieli, $X_{n_2\times n_2}^{(4)}$ - macierz ocen nauczycieli przez nauczycieli. W praktyce ocenianie jednych studentów lub nauczycieli przez drugich mogłoby prowadzić do niezręcznych sytuacji i konfliktów międzyludzkich. Gdyby przyjąć, że nie dopuszcza się jakiejkolwiek oceny studentów przez innych studentów oraz nauczycieli przez innych nauczycieli, to macierze $X_{n_1\times n_1}^{(1)}$ i $X_{n_2\times n_2}^{(4)}$ zerowałyby się i macierz (3.1) przybrałaby postać

$$\begin{pmatrix}
O & X_{n_1\times n_2}^{(2)} \\
X_{n_2\times n_1}^{(3)} & O
\end{pmatrix}, \qquad (3.3)$$

czyli

$$\begin{pmatrix} 0 & 0 & \dots & 0 & x^{(2)}_{1,1} & x^{(2)}_{1,2} & \dots & x^{(2)}_{1,n_2} \\ 0 & 0 & \dots & 0 & x^{(2)}_{2,1} & x^{(2)}_{2,2} & \dots & x^{(2)}_{2,n_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & x^{(2)}_{n_1,1} & x^{(2)}_{n_1,2} & \dots & x^{(2)}_{n_1,n_2} \\ x^{(3)}_{1,1} & x^{(3)}_{1,2} & \dots & x^{(3)}_{1,n_1} & 0 & 0 & \dots & 0 \\ x^{(3)}_{2,1} & x^{(3)}_{2,2} & \dots & x^{(3)}_{2,n_1} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{(3)}_{n_2,1} & x^{(3)}_{n_2,2} & \dots & x^{(3)}_{n_2,n_1} & 0 & 0 & \dots & 0 \end{pmatrix}. \qquad (3.4)$$

Jeśli przyjąć, że żaden student (nauczyciel) nie ocenia innych studentów (nauczycieli) prócz samego siebie, to $X^{(1)S}_{n_1\times n_1}$ i $X^{(4)}_{n_2\times n_1}$ stają się macierzami diagonalnymi i macierz (3.1) przyjmuje postać

$$\begin{pmatrix} x^{(1)}_{1,1} & 0 & \dots & 0 & x^{(2)}_{1,1} & x^{(2)}_{1,2} & \dots & x^{(2)}_{1,n_2} \\ 0 & x^{(1)}_{2,2} & \dots & 0 & x^{(2)}_{2,1} & x^{(2)}_{2,2} & \dots & x^{(2)}_{2,n_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x^{(1)}_{n_1,n_1} & x^{(2)}_{n_1,1} & x^{(2)}_{n_1,2} & \dots & x^{(2)}_{n_1,n_1} \\ x^{(3)}_{1,1} & x^{(3)}_{1,2} & \dots & x^{(3)}_{1,n_1} & x^{(4)}_{1,1} & 0 & \dots & 0 \\ x^{(3)}_{2,1} & x^{(3)}_{2,2} & \dots & x^{(3)}_{2,n_1} & 0 & x^{(4)}_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x^{(3)}_{n_2,1} & x^{(3)}_{n_2,2} & \dots & x^{(3)}_{n_2,n_1} & 0 & 0 & \dots & x^{(4)}_{n_2,n_2} \end{pmatrix}, \qquad (3.5)$$

gdzie elementy $x^{(1)}_{i,j}$ $(i,j=1,2,..,n_1)$ są samoocenami studentów, zaś elementy $x^{(4)}_{i,j}$ $(i,j=1,2,...,n_2)$ są samoocenami nauczycieli.

Ze względów racjonalnych przyjmujemy, że macierz ocen ma postać (3.5). Jest to równoważne postulowaniu, że każdy student jest oceniany przez każdego nauczyciela, zaś każdy nauczyciel może być oceniony przez każdego studenta, przy czym każdy każdy student i każdy nauczyciel dokonują samooceny, ale nie oceniają swych kolegów. Oczywiście, niedopuszczalna jest sytuacja, by student oceniał nauczyciela lub nauczyciel oceniał studenta, jeśli nie ma między nimi żadnego bezpośredniego kontaktu edukacyjnego. Przyjmujemy, że w tym przypadku wartości ocen wystawianych przez studentów nauczycielom bądź przez nauczycieli studentom są równe zeru.

Wchodzą w grę dwie możliwości: oceny mogą być robione z punktu widzenia jednej z góry ustalonej cechy (np. pomocą oceny z matematyki na świadectwie maturalnym, średniej ocen z kilku wybranych przedmiotów) lub z punktu widzenia skończonej liczby z góry ustalonych cech (np. za pomocą ocen z matematyki, fizyki i informatyki). W pierwszym przypadku elementami macierzy (3.1)-(3.5) są skalary, w drugim – wektory. Pierwszy przypadek będziemy nazywać skalarnym lub jednowymiarowym, drugi – wektorowym lub wielowymiarowym. Ponieważ sytuacja skalarna jest szczególnym przypadkiem wektorowej, więc większy nacisk będziemy kłaść na ocenę wielowymiarową.

Względy racjonalności wymagają, by macierz ocen traktować jako macierz niepewną. Możliwe są różne rodzaje niepewności i różne ich reprezentacje matematyczne (Miller i Starr 1969). W ślad za Guilfordem (1964) zakładamy, że mamy do czynienia z niepewnością probabilistyczno-statystyczną, tzn. że elementy macierzy ocen są realizacjami zmiennych losowych (skalarnych lub wektorowych) o znanych lub nieznanych rozkładach prawdopodobieństwa. Zakładamy też, że zmienne te przyjmują wartości liczbowe zgodne z zakresami zmienności skal odpowiadających poszczególnym kryteriom. Najczęściej stosuje się skale nominalne, porządkowe, przedziałowe lub ilorazowe.

Podstawowym warunkiem poddania zbioru danych (3.5) jakiejkolwiek obróbce statystycznej jest jego statystyczna jednorodność w sensie rozpatrywanych kryteriów oceny. W statystyce matematycznej od dawna rozwijają się metody weryfikacji hipotezy o jednorodności zbioru danych, przede wszystkim jednowymiarowych. Metody te pozwalają wprawdzie rozstrzygnąć, czy zbiór jest jednorodny czy niejednorodny, ale – w przypadku stwierdzenia niejednorodności – nie wyznaczają podziału zbioru niejednorodnego na podzbiory jednorodne. Jeszcze jaskrawiej brak ten ujawnia się w przypadku wielowymiarowych zbiorów danych. Nie tylko w kontekście edukacji istnieje więc potrzeba konstruowania takich metod weryfikacji jednorodności wielowymiarowych zbiorów danych, które – gdy zbiór okaże się niejednorodny – wyznaczą jednoznaczny jego podział na jak najmniejszą liczbę podzbiorów jednorodnych.

4. Zastosowanie statystycznych metod badania jednorodności wielowymiarowych zbiorów danych do analizy procesów edukacyjnych

4.1. Uwagi wstępne

Jednym z podstawowych przedsięwzięć organizacyjnych w szkole wyższej jest podział zbiorowości studentów rozpoczynających naukę na grupy edukacyjne, które niejednokrotnie nie są zmieniane przez cały okres studiów. Podział ten jest robiony w sposób administracyjny, więc do tak tworzonych grup trafiają studenci mniej lub bardziej zróżnicowani pod względem merytorycznego przygotowania do studiów. Utrudnia to, a czasem wręcz uniemożliwia, prowadzenie zajęć – w takiej bowiem sytuacji poziom zajęć jest albo zbyt wysoki dla studentów słabszych, albo zbyt niski dla dobrych. Każda grupa studencka powinna być więc być merytorycznie jednorodna, tzn. poziom znajdujących się w niej studentów powinien być mniej więcej wyrównany. Zachodzi więc potrzeba sprawdzania stopnia jednorodności grup utworzonych w sposób administracyjny i ewentualnego wprowadzania odpowiednich korekt.

U podstaw administracyjnego sposobu tworzenia grup studenckich leży założenie, że liczba studentów w każdej grupie nie może być większa od pewnej z góry ustalonej liczby całkowitej – najczęściej odpowiadającej średniej liczbie miejsc w sali zajęciowej – ani też mniejsza od minimalnej liczebności gwarantującej ekonomiczną opłacalność prowadzenia zajęć. Tak więc liczba grup jest z góry znana, a przydziału studentów do każdej z nich dokonuje się w sposób mechaniczny.

W dotychczasowych badaniach – zgodnie z koncepcją Guilforda (1964) – z góry zakłada się, że oceny studentów i nauczycieli mają charakter probabilistyczno-statystyczny. Zapomina się jednak, że przyjęcie tego założenia jest możliwe jedynie wtedy, gdy przyjmie się, iż zbiorowości studentów i nauczycieli (które charakteryzuje się za pomocą jednowymiarowych lub wielowymiarowych zbiorów danych – ocen) są populacjami statystytycznymi w sensie jaki temu pojęciu nadaje statystyka matematyczna. Aby więc móc prowadzić wnioskowanie statystyczne trzeba najprzód sprawdzić statystyczną jednorodność każdej z tych zbiorowości i – w razie stwierdzenia niejednorodności – podzielić ją na zbiorowości statystycznie jednorodne, które będą mogły być poddane dalszej obróbce. Przyjęcie a priori założenia o statystycznym charakterze zbiorowości studentów lub nauczycieli jest błędem metodologicznym, którego konsekwencją jest nadużywanie metody statystycznej badaniach z zakresu edukacji. Kubik (1998) pisze: „Przypuśćmy, że prowadzimy badania nad przyczynami niepowodzeń w szkole (badania takie są rzeczywiście prowadzone przez instytuty naukowe). W tym celu badamy sprawę zależności między wynikami w nauce a wszelkimi potencjalnymi „przyczynami" (liczbą rodzeństwa. odległością szkoły od domu, wykształceniem nauczycieli, ...). Robimy to w ten sposób, że stawiamy kolejne hipotezy H_i $(i = 1,2,...)$ o niezależności wyników w nauce od i-tej z kolei potencjalnej „przyczyny". Dla każdej z tych hipotez stosujemy test chi-kwadrat na poziomie istotności 0.05, opierając się na zebranych danych statystycznych. Należy oczekiwać, że choćby wszystkie hipotezy były prawdziwe, to prędzej czy później odrzucimy którąś z nich (odrzucenie hipotezy prawdziwej zdarza się rzadko, ale jednak się zdarza). Odrzucenie hipotezy H_j uważane jest za wykrycie zależności pomiędzy wynikami w nauce a j-tą przyczyną. W ten sposób można np. „odkryć" zależność pomiędzy wynikami w nauce a noszeniem teczki lub tornistra." Kontynuując tę myśl Kubik stwierdza: „Podany wyżej sposób „wykrywania nowych faktów naukowych" jest humorystyczny, jednakże bywa on stosowany. Zostało to już dawno zauważone i skrytykowane. W szczególności pisał o tym światowej sławy probabilista amerykański W. Feller (1969). Publikowanie „istotnych" wyników otrzymanych w wyżej opisany sposób Feler nazywa skandalem. Sądzę, że w ciągu ponad ćwierć wieku, które minęły od ukazania się artykułu Fellera, skandale te nie zniknęły, ale przeciwnie – nasiliły się. Dzięki powszechności stosownia komputerów „wykrywanie faktów naukowych" w opisany sposób staje się bowiem coraz łatwiejsze". Opinia Fellera, na którą powołuje się Kubik, nie jest odosobniona. Na fakt nadużywania metody wnioskowania statystycznego zwracają uwagę probabiliści tej miary, co Tatubalin (1977), Barra (1982), Rao

(1994), Bartoszyński i Niewiadomska-Bugaj (1996) i inni. Trudno nie zgodzić się z ich opinią, że zastosowania statystyki matematycznej stały się polem wielkich nadużyć naukowych, wynikających – przede wszystkim – z niezrozumienia istoty podejścia statystycznego. Do najczęściej popełnianych błędów należy traktowanie a priori wszelkich zbiorów danych jako zbiorowości statystycznych i – w oparciu o te zbiory – estymowanie nieznanych wartości parametrów modeli matematycznych lub weryfikowanie rozmaitych hipotez dotyczących własności modeli za pomocą wnioskowania statystycznego.

Jeżeli dana zbiorowość spełnia warunki uprawniające do uznania jej za zbiorowość statystyczną, to jednym z podstawowych wymogów uprawniających do prowadzenia w oparciu o nią wnioskowania statystycznego jest jej jednorodność. Zanim więc zbiór danych (np. zbiór ocen charakteryzujących merytoryczną jakość studentów lub nauczycieli) poddamy obróbce statystycznejm, trzeba najprzód sprawdzić, czy jest on jednorodny. O jednorodności zbiorowości można jednak mówić tylko i wyłącznie w kontekście pewnego z góry określonego kryterium jednorodności. Postać kryterium powinna być zgodna z naturą rozpatrywanego zbioru danych oraz z metodologią rachunku prawdopodobieństwa i statystyki matematycznej. Ponieważ centralne miejsce w tej metodologii zajmuje rozkład normalny, z reguły konstruuje się kryteria oparte na hipotezie, że rozkład badanej cechy statystycznej w populacji generalnej jest normalny. Nie oznacza to bynajmniej przyjęcia, że rzeczywiste rozkłady zmiennych losowych też muszą być normalne. Im bliższe są one jednak rozkładowi normalnemu, tym lepsza jest zgodność teoretycznej wartości kryterium z jego wartością empiryczną.

Historycznie najstarszą a zarazem najprostszą metodą badania jednorodności jednowymiarowego zbioru danych statystycznych jest metoda zaproponowana przez Abbego (1863), udoskonalona następnie przez Helmerta (1876). Obecnie jest ona nazywana metodą Abbego-Helmerta. Omówimy istotę tej metody i podstawowe wyniki dotyczące własności stosowanej w niej statystyki, otrzymane przede wszystkim przez von Neumanna (1941, 1942). Uogólnimy tę statystykę na przypadek zbiorowości wielowymiarowej. Zaproponujemy metodę wykorzystania uogólnionej wersji statystyki Abbego-Helmerta do badania jednorodności wielowymiarowej zbiorowości z punktu widzenia wielu cech równocześnie. Metodę tę przedstawimy w dwóch wersjach. Pierwsza będzie oparta na idei weryfikacji jednorodności badanej zbiorowości i – w razie stwierdzenia niejednorodności – podziale zbiorowości na dwie części, a następnie badaniu jednorodności każdej z nich itd., aż do wyczerpania możliwości dalszego podziału. Druga będzie oparta na założeniu, że rozpatrywana zbiorowość jest z góry podzielona na pewną liczbę części, których jednorodność – jako elementów tej zbiorowości – musi być sprawdzona dla wszystkich zbiorów równocześnie. Rozważania teoretyczne zostaną zilustrowane przykładem zastosowania metody. W rozważaniach akcent będzie położony na przypadek zbiorowości wielowymiarowej. Formuły dla zbiorowości jednowymiarowej będą w zasadzie podane jako szczególny przypadek formuł wielowymiarowych.

4.2. Metoda Abbego-Helmerta

4.2.1. Idea metody - statystyka Q

Abbe (1863) i Helmert (1876) rozpatrywali zbiór danych $x_1, x_2, ..., x_n$, o których założyli, że są realizacjami jednowymiarowych niezależnych zmiennych losowych normalnych $X_1, X_2, ..., X_n$ o takich samych, ale nieznanych wariancjach. Zaproponowali, by uznać, że zbiór jest statystycznie jednorodny, jeśli wartości oczekiwane zmiennych X_i $(i = 1,2,...,n)$ są takie same. W tym przypadku hipoteza zerowa H_0 ma postać

$$H_0 : E[X_1] = E[X_2] = ... = E[X_n], \tag{4.1}$$

gdzie E jest operatorem wartości oczekiwanej. Za hipotezę alternatywną Abbe i Helmert uznali

$$H_i : | E(X_{i+1}) - E(X_i) | > 0, \quad (i = 1,2,...,n-1). \tag{4.2}$$

Jako kryterium weryfikacji hipotezy zerowej przyjęli połowę stosunku średniej kwadratowej kolejnych odchyleń obserwacji,

$$\delta^2 = \frac{1}{n} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2, \tag{4.3}$$

do ich wariancji

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{4.4}$$

gdzie

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{4.5}$$

czyli statystykę

$$Q = \frac{1}{2} \frac{\delta^2}{s^2}. \tag{4.6}$$

Hipoteza H_0 o równości wartości oczekiwanych zmiennych losowych X_i $(i = 1,2,...,n)$ (a więc o statystycznej jednorodności rozpatrywanego zbioru danych) zostaje zaakceptowana, jeśli empiryczna wartość statystyki Q jest niemniejsza od wartości krytycznej $Q_n(\alpha)$, gdzie n jest licznością zbioru obserwacji a α jest z góry przyjętym poziomem istotności. Wartość krytyczna $Q_n(\alpha)$ jest kwantylem rzędu α rozkładu zmiennej losowej Q, czyli rozwiązaniem równania $P\{Q < Q_n(\alpha)\} = \alpha$. Hipotezę zerową odrzuca się, jeśli empiryczna wartość statystyki Q jest mniejsza od wartości krytycznej $Q_n(\alpha)$.

4.2.2. Podstawowe własności statystyki Q

Statystyka Q, jako funkcja zmiennych losowych X_i $(i = 1,2,\dots,n)$, jest też zmienną losową. Badaniem własności tej zmiennej zajmowało się wielu naukowców (von Neumann 1941; Williams 1941; Young 1941; von Neumann, Kent, Bellinson i Hart 1941; Hart 1942; von Neumann 1942: Bingham i Nelson 1981; Alexandersson 1986). Podstawowe wyniki, leżące u podstaw wszystkich dalszych badań, podał von Neumann (1941). Zamiast formuły (4.3) przyjął on estymator

$$\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \tag{4.7}$$

i rozpatrywał statystykę $\eta = 2Q$, czyli

$$\eta = \frac{\delta^2}{s^2}. \tag{4.8}$$

Przekształcił ją do postaci

$$\eta = \frac{2n}{n-1}(1-\varepsilon), \tag{4.9}$$

gdzie

$$\varepsilon = 1 - \frac{1}{2}Q. \tag{4.10}$$

i udowodnił, że zmienna losowa ε daje się przedstawić w formie następującego szeregu

$$\varepsilon = \sum_{i=1}^{n-1}\cos\frac{i\pi}{n}\cdot x_i^2, \tag{4.11}$$

przy czym punkty $x_1, x_2, \dots, x_{n-1}$ mają rozkład jednostajny na sferze o równaniu

$$\sum_{i=1}^{n-1} x_i^2 = 1 . \tag{4.12}$$

Von Neumann wykazał też, że rozkład zmiennej losowej ε jest symetryczny względem punktu 0, wskutek czego

$$E(\varepsilon) = 0 \tag{4.13}$$

oraz że jej wariancja wynosi

$$D^2(\varepsilon) = 0 \tag{4.14}$$

(tzn. że ma ona zerową wartość oczekiwaną). Udowodnił ponadto, że maksymalną wartością statystyki ε jest $\cos\frac{\pi}{n}$, zaś minimalną $\cos\frac{(n-1)\pi}{n} = -\cos\frac{\pi}{n}$. Wyprowadził formuły na obliczanie momentów centralonych rozkładu statystyki ε i wykazał, że wszystkie momenty rzędu nieparzystego są równe zeru, natomiast momenty rzędu parzystego mają wartości niezerowe, przy czym, w szczególności,

$$D^2(\varepsilon) = \frac{n-2}{n^2-1}, \tag{4.15}$$

Z wyników von Neumanna bezpośrednio wnioskuje się, że rozkład statystyki η jest symetryczny względem jej wartości średniej, którą jest $\frac{2n}{n-1}$, oraz że maksymalną wartością statystyki η jest $\frac{2n}{n-1}\left(1+\cos\frac{\pi}{n}\right) = \frac{4n}{n-1}\cos^2\frac{\pi}{2n}$, zaś minimalną $\frac{2n}{n-1}\left(1-\cos\frac{\pi}{n}\right) = \frac{4n}{n-1}\sin^2\frac{\pi}{2n}$. W odniesieniu do statystyki Q wnioskuje się, że jej wartość oczekiwana i wariancja wyrażają się – odpowiednio – formułami

$$E(Q) = 1 , \tag{4.16}$$

$$D^2(Q) = \frac{n-2}{n^2-1} . \tag{4.17}$$

Ważnym wynikiem otrzymanym przez von Neumanna (1941) było wykazanie, że gdy $n \to \infty$, wtedy rozkład statystyki ε staje się rozkładem asymptotycznie normalnym o zerowej wartości oczekiwanej oraz odchyleniu standardowym $\sqrt{\frac{1}{n}}$

i że – wobec tego – dla dużych wartości n, można rozkład statystyki Q można aproksymować za pomocą rozkładu zmiennej losowej

$$1+\frac{X}{\sqrt{n+0.5(1+X^2)}}, \tag{4.18}$$

gdzie X jest zmienną gaussowską o zerowej wartości oczekiwanej i jednostkowej wariancji.

Ciekawym kierunkiem badań jest poszukiwanie przybliżonych metod obliczania wartości wariancji statystyki Q. Jednym z bardziej istotnych wyników jest następująca formuła zaproponowana przez Harta (1942), Halda (1952) oraz Linnika (1962)

$$D^2(Q)\approx\frac{1}{n+1}. \tag{4.19}$$

Zwrócili oni uwagę, że formuła ta może być stosowana tylko dla dużych wartości n, przy czym najlepszą zgodność z rzeczywistością otrzymuje się, gdy $n>60$. Udowodnili też, że przy korzystaniu z formuły (4.19) błąd obliczeń jest rzędu $O\left(\frac{1}{n^2}\right)$.

4.3. Metody oparte na relacji równoważności zmiennych losowych

4.3.1. Zmienne losowe równoważne. Ogólne sformułowanie warunków jednorodności

Rozpatrzmy skończony, n-elementowy zbiór studentów (nauczycieli) $S=\{s_k : k=1,2,...,n\}$. Przyjmijmy, że studentowi s_k $(k=1,2,...,n)$ jest przyporządkowana zmienna losowa X_k o dystrybuancie $F_k(X)$. Niech X^k będzie zbiorem (wektorem) wszystkich zmiennych losowych X_k, tzn. $X^k=(X_1,X_2,...,X_n)$. Rozpatrzmy zmienne losowe $X_{k_1},X_{k_2}\in X^k$. Powiemy, że są one sobie równoważne, jeśli mają taki sam rozkład prawdopodobieństwa, tzn. jeśli $F_{k_1}(X)=F_{k_2}(X)$. Oznaczmy zbiór wartości argumentu X przez X'.

Jak wiadomo (Rasiowa 1979), każda relacja równoważności $\approx$ w zbiorze niepustym Z wyznacza podział tego zbioru na rozłączne i niepuste podzbiory, mianowicie na klasy równoważności tej relacji, w taki sposób, że dwa elementy $z', z'' \in Z$ należą do jednej i tej samej klasy równoważności wtedy i tylko wtedy, gdy $z' \approx z''$. Skoro tak, to wprowadzona przez nas relacja równoważności zmiennych losowych w zbiorze X^k wyznacza podział tego zbioru na klasy równoważności. Podziałowi temu odpowiada następująca dekompozycja zbioru S na podzbiory

$$S = S_1 \cup S_2 \cup \ldots \cup S_q , \quad (q \geq 1) . \tag{4.20}$$

Jeśli $q = 1$, to zbiór S jest statystycznie jednorodny. W przypadku, gdy $q \geq 1$, zbiór S jest statystycznie niejednorodny.

Wprowadzone wyżej pojęcie równoważności zmiennych losowych można wykorzystać do zdefiniowania pojęcia statystycznej jednorodności zbioru. Powiemy, że zbiór $X^R \subset X^k$ jest statystycznie jednorodny, jeśli dla każdej pary elementów $k', k'' \in R$ i dla każdego $X \in X'$ zachodzi równość

$$F_{k'}(X) - F_{k''}(X) = 0 . \tag{4.21}$$

Z formuły tej logicznie wynikają dwa następujące wnioski. Po pierwsze, klasa $X^{R'}$ zbiorów X^R jest klasą zbiorów jednorodnych, jeśli dla każdego $R \in R'$ i dla każdego $X \in X'$ zachodzi równość

$$F_{k'} \in R(X) - F_{k''} \in R(X) = 0 . \tag{4.22}$$

Po drugie, rodzina zbiorów jednorodnych jest zbiorem jednorodnym, jeśli dla każdej pary $R_1, R_2 \in R'$ i dla każdego , . , .

$$F_k \in R_1(X) - F_k \in R_2(X) = 0 . \tag{4.23}$$

A zatem, jeśli dla zbioru X^R jest spełniona równość (4.21) i jeśli zbiór ten pokrywa się ze zbiorem X^k, to zbiór X^k jest jednorodny.

Mając określone warunki statystycznej jednorodności zbioru można – przez ich zaprzeczenie – sformułować warunki niejednorodności. Jeśli więc w podzbiorze X^R zbioru X^k istnieje taka para elementów $k', k'' \in R$, że dla co najmniej jednego $X \in X'$

$$| F_{k'}(X) - F_{k''}(X) | > 0 , \tag{4.24}$$

to podzbiór Ξ^R jest statystycznie niejednorodny. Stąd wynika, że jeśli w klasie zbiorów jednorodnych istnieje choćby tylko jedna para zbiorów X^{R_1}, X^{R_2} takich, że dla co najmniej jednego $X \in X'$

$$|F_k \in R_1(X) - F_k \in R_2(X)| > 0, \tag{4.25}$$

to rozpatrywany zbiór jest niejednorodny. Co więcej, jeśli warunek (4.25) jest spełniony dla każdej pary zbiorów $X^{R_1}, X^{R_2} \in X^k$, to cały zbiór wyjściowy jest niejednorodny. Jeśli jednak zbiór niejednorodny zawiera co najmniej jedną parę podzbiorów, dla której jest spełniony warunek (4.22), to zbiór ten jest tylko częściowo niejednorodny.

4.3.2. Warunki jednorodności – przypadek jednowymiarowy

Kryteria (4.21)–(4.25) statystycznej jednorodności lub niejednorodności zbioru danych mają charakter ogólny – w toku ich formułowania nie robiliśmy żadnych założeń co do postaci rozkładu zmiennych losowych X_{kj} $(k = 1,2,...,n;\ j = 1,2,...,m)$. Przyjmijmy teraz, że są to zmienne gaussowskie (zmienne niegaussowskie można w wielu przypadkach przekształcić za pomocą odpowiedniej transformacji matematycznej w zmienne gaussowskie). Oczywiście nie należy z tego wnioskować, że empiryczne rozkłady tych zmiennych też muszą być gaussowskie. Wiadomo bowiem (zob., np., Oktaba 1966; Fisz 1969; Hellwig 1980), że stopień zgodności rozkładu teoretycznego z empirycznym ocenia się testując hipotezę o normalności rozkładu statystyki, odpowiadającej użytemu kryterium zgodności. Ostateczną decyzję o przyjęciu lub odrzuceniu hipotezy podejmuje się w oparciu o zasadę praktycznej pewności (Cramér 1945), której formalnym wyrazem jest poziom istotności hipotezy, nazywany też współczynnikiem istotności.

Zakładamy więc, że wszystkie jednowymiarowe zmienne losowe ξ_{kj} $(k = 1,2,...,n;\ j = 1,2,...,m)$ mają rozkłady normalne. Skoro tak, to zmienna losowa X_k ma m-wymiarowy rozkład normalny, określony wzorem

$$F(X;\Theta_k,V_k) = (2\pi)^{-\frac{m}{2}} |V_k|^{-\frac{m}{2}} \exp\left[-\frac{1}{2}(X-\Theta_k)'V_k^{-1}(X-\Theta_k)\right], \tag{4.26}$$

gdzie $\Theta_k = (\Theta_{k1},\Theta_{k2},...,\Theta_{km})$ jest wektorem wartości oczekiwanych zmiennych ξ_{kj}, V_k jest macierzą kowariancji (macierz kwadratowa stopnia m), $|V_k|$ – jej wyznacznikiem, a symbol $(X-\Theta_k)'$ oznacza transpozycję macierzy $(X-\Theta_k)$.

Wykorzystując formułę (4.23) można warunek statystycznej jednorodności (4.21) zapisać w postaci

$$F(X;\Theta_{k'},V_{k'})-F(X;\Theta_{k''},V_{k''})=0\,, \tag{4.27}$$

przy czym równość ta musi zachodzić dla każdego $X \in X'$. Zauważmy, że równość (4.27) jest równoważna następującemu układowi warunków

$$\Theta_{k'}-\Theta_{k''}=\underbrace{(0,0,...,0)}_{m}\,,\qquad V_{k'}=V_{k''}\,,\qquad (k',k''\in R)\,. \tag{4.28}$$

Warunki te pozwalają przedstawić kryterium niejednorodności (4.24) w postaci nierówności

$$|\,F(X;\Theta_{k'},V_{k'})-F(X;\Theta_{k''},V_{k''})\,|>0\,, \tag{4.29}$$

która musi zachodzić dla co najmniej jednego $X \in X'$, lub w postaci warunku

$$\Theta_{k'}-\Theta_{k''}\neq\underbrace{(0,0,...,0)}_{m}\,, \tag{4.30}$$

który musi być spełniony dla co najmniej jednej pary elementów $k',k''\in S$, jeśli dla wszystkich par $k',k''\in S$ zachodzi równość $V_{k'}=V_{k''}$.

Podobne rozumowanie prowadzi do stwierdzenia, że w przypadku m-wymiarowego rozkładu normalnego (zob. (4.26)) formuły (4.22) i (4.24) przyjmą – odpowiednio – postaci:

1. Jeśli zbiór X^k jest jednorodny, to dla każdej pary X^{R_l}, X^{R_k} i dla wszystkich $R_l,R_k\in S$ zachodzi równość

$$\frac{1}{n_l}\sum_{k\in R_l}\Theta_k-\frac{1}{n_k}\sum_{k\in R_k}\Theta_k=\underbrace{(0,0,...,0)}_{m}\,. \tag{4.31}$$

2. Jeśli zbiór X^k jest niejednorodny, to zbiór P^2 wszystkich możliwych podziałów zbioru S na dwie części zawiera co najmniej jeden taki element, że

$$\frac{1}{n_1}\sum_{k\in R_1}\Theta_k-\frac{1}{n_2}\sum_{k\in R_2}\Theta_k\neq\underbrace{(0,0,...,0)}_{m}\,. \tag{4.32}$$

Jeżeli przyjąć, że dla $k = 1,2,\dots,n$ zmienne losowe X_{kj} $(j = 1,2,\dots,m)$ – a więc składowe wektora X_k – są niezależne, zaś dla $j = 1,2,\dots,m$ zmienne losowe X_{kj} $(k = 1,2,\dots,n)$ mają jednakowe wariancje, to wyrażeniom (4.28) i (4.32) można nadać postać

$$\delta(\rho^2) = \frac{1}{n_1}\sum_{k\in R_1}\Theta_k - \frac{1}{n_2}\sum_{k\in R_2}\Theta_k\,, \tag{4.33}$$

gdzie ρ^2 jest podzbiorem podziałów zbiorowości statystycznej na całym zbiorze P^2. Funkcja $\delta(\rho^2)$ jest kryterium statystycznej jednorodności m-wymiarowej zbiorowości danych.

Założenia, które ograniczają swobodę korzystania z wyrażenia (4.33), są wprawdzie dość silne, lecz mimo to w wielu sytuacjach rzeczywistych w pełni dopuszczalne. Na przykład, często zdarza się, że w ramach każdej odrębnie rozpatrywanej grupy danych statystycznych wartość współczynnika korelacji jest bliska zeru, podczas gdy w pełnej zbiorowości danych ma miejsce silna korelacja. W takich przypadkach można rozszerzyć obserwowaną niezależność zmiennych losowych w odrębnych grupach na analizowane zmienne losowe X_{kj} $(j = 1,2,\dots,m)$. Postulat o równości wariancji też jest uzasadniony, jeśli rozpatruje się małe podzbiory, dla których jest możliwe zachowanie warunku $X_{11} = X_{22} = \dots = X_{nm}$, co jest równoważne równości wariancji zmiennych losowych jednowymiarowych wedle cechy j $(j = 1,2,\dots,m)$.

W dziedzinie edukacji, a zwłaszcza w sferze ewaluacji procesu edukacyjnego, większość danych ma charakter wielowymiarowy. Przejdziemy więc teraz do omówienia warunków jednorodności zbiorow danych wielowymiarowych.

4.3.3. Warunki jednorodności – przypadek wielowymiarowy

Z formuły (4.33) wynika, że hipoteza o jednorodności jest prawdziwa, jeśli wartość m-wymiarowej funkcji $\delta(\rho^2)$ jest równa 0. Przeanalizujmy, jaki jest sens tej hipotezy w przypadku, gdy zbiór S jest wielowymiarowy.

Tak jak poprzednio, rozpatrujemy skończoną zbiorowość studentów $S = \{s_k : k = 1,2,\dots,n\}$. Zakładamy, że student s_k $(k = 1,2,\dots,n)$ jest opisany za pomocą m-wymiarowej zmiennej losowej $X_k = (X_{k1}, X_{k2},\dots,X_{km})$, gdzie X_{kj} $(k = 1,2,\dots,n)$ – zmienna losowa jednowymiarowa charakteryzująca go z punktu widzenia cechy j $(j = 1,2,\dots,m)$. Cała zbiorowość studentów jest więc scharakteryzowana za pomocą takiej macierzy losowej

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{pmatrix}. \tag{4.34}$$

Oznaczmy przez

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \tag{4.35}$$

macierz, której elementami są zaobserwowane wartości zmiennych losowych X_{kj} $(k = 1,2,\dots,n;\ j = 1,2,\dots,m)$. Zauważmy, że wiersze tej macierzy są realizacjami m-wymiarowych zmiennych losowych $X_k = (X_{k1}, X_{k2}, \dots, X_{km})$, gdzie $(s = 1,2,\dots,n)$.

Zakładamy, że zmienne losowe $X_k = (X_{k1}, X_{k2}, \dots, X_{km})$ są normalne, tzn. odnosi się do nich formuła (4.26). Dla zachowania płynności rozumowania napiszmy ją tu jeszcze raz:

$$F(X_k) = (2\pi)^{-\frac{m}{2}} \,|\, V_k \,|^{-\frac{m}{2}} \exp\left[-\frac{1}{2}(X_k - \Theta_k)' V_k^{-1}(X_k - \Theta_k) \right], \tag{4.36}$$

gdzie $\Theta_k = (\Theta_{k1}, \Theta_{k2}, \dots, \Theta_{km})$ jest wektorem wartości oczekiwanych zmiennych X_{kj}, V_k jest macierzą kowariancji tych zmiennych, tj.

$$V_k = \begin{pmatrix} \sigma_{11}^k & \sigma_{12}^k & \dots & \sigma_{1m}^k \\ \sigma_{21}^k & \sigma_{22}^k & \dots & \sigma_{2m}^k \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^k & \sigma_{m2}^k & \dots & \sigma_{mm}^k \end{pmatrix}, \quad (k = 1,2,\dots,n) \tag{4.37}$$

przy czym $\sigma_{hl}^s = E[(X_h - E(X_h))(X_l - E(X_l))]$ dla $h,l = 1,2,\dots,m$, zaś $|\, V_k \,|$ jest wyznacznikiem macierzy V_k.

Jeżeli zmienne losowe $X_{s1}, X_{s2}, \dots, X_{sm}$ są niezależne, to ich współczynniki kowariancji są równe zeru i V_k staje się macierzą diagonalną

$$V_k = \begin{pmatrix} \sigma_{11}^k & 0 & \dots & 0 \\ 0 & \sigma_{22}^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{mm}^k \end{pmatrix}. \qquad (4.38)$$

Ponieważ $\sigma_{hh}^k = E[(X_h^k - E(X_h^k))(X_h^k - E(X_h^k))] = (\sigma_h^k)^2 = \sigma_{kh}^2$, więc elementami przekątniowymi tej macierzy są wariancje kolejnych zmiennych losowych. W tym przypadku forma kwadratowa $(X_k - \Theta_k)'V_k^{-1}(X_k - \Theta_k)$, która występuje w formule (4.36) w wykładniku funkcji ekspotencjalnej, przyjmuje postać

$$(X_k - \Theta_k)'V_k^{-1}(X_k - \Theta_k) = \sum_{j=1}^{m} \frac{(x_j - \Theta_{kj})^2}{\sigma_{ij}}, \qquad (4.39)$$

zaś wyznacznik macierzy kowariancji jest równy iloczynowi jej elementów diagonalnych

$$|V_k| = \prod_{h=1}^{m} \sigma_{kh}^2. \qquad (4.40)$$

Wobec tego, w przypadku niezależności zmiennych losowych $X_{k1}, X_{k2}, \dots, X_{km}$, funkcja (4.36) przyjmuje postać

$$F(X_k) = (2\pi)^{-\frac{m}{2}} \left(\prod_{h=1}^{m} \sigma_h \right)^{-m} \exp\left[-\frac{1}{2} \sum_{h=1}^{m} \frac{(x_h - \Theta_h)^2}{\sigma_k^2} \right]. \qquad (4.41)$$

W dalszych rozważaniach przyjmujemy, że $X_k = (X_{k1}, X_{k2}, \dots, X_{km})$ są zmiennymi losowymi niezależnymi, których wartości oczekiwane tworzą wektor $\Theta_k = (\Theta_{k1}, \Theta_{k2}, \dots, \Theta_{km})$, mającymi taką samą diagonalną macierz kowariancji. Przyjęcie tych założeń pozwala związać z hipotezą o jednorodności zbioru następujące równości

$$\Theta_{s_k} - \Theta_{s_l} = \underbrace{(0,0,\dots,0)}_{m}, \quad \text{dla wszystkich} \quad s_k, s_l \in S, \qquad (4.42)$$

$$\delta(\rho^2) = \underbrace{(0,0,\dots,0)}_{m}, \quad \text{dla wszystkich} \quad \rho^2 \in P^2, \qquad (4.43)$$

gdzie ρ^2 jest podzbiorem podziałów rozpatrywanej zbiorowości.

Słabą stroną kryterium (4.42) jest niemożność zweryfikowania hipotezy, jeśli każdemu $s \in S$ odpowiada tylko jedna obserwacja. W tym przypadku kryterium to jest równoważne układowi $\frac{n(n-1)}{2}$ równości $\Theta_{s'} - \Theta_{s''} = \underbrace{(0,0,...,0)}_{m}$. Jeśli jednak $n = 1$ (zbiór jednoelementowy), to równości takie po prostu nie istnieją. A zatem, kryterium (4.42) zasadniczo nie nadaje się do weryfikacji hipotezy o jednorodności zbioru złożonego z n obserwacji ($n > s$, $i = 1,2,...,s$). Znacznie wygodniej jest korzystać z kryterium (4.43), ponieważ funkcja $\delta(\rho^2) = \underbrace{(0,0,...,0)}_{m}$ jest określona na zbiorze podziałów zbioru X^s na dwie rozłączne części X^{R_1} i X^{R_2}, a więc weryfikacja hipotezy o jednorodności za pomocą wyrażenia (4.40) pozwala dla każdego $\rho^2 \in P^2$ brać pod uwagę pełen zbiór n obserwacji.

Hipotezą alternatywną będzie zbiór alternatyw H_1 określonych za pomocą nierówności

$$H_1 : \delta(\rho^2) \neq \underbrace{(0,0,...,0)}_{m} \tag{4.44}$$

dla co najmniej jednego podziału $\rho^2 \in P^2$.

Podzielmy macierz (4.34) na dwie rozłączne podmacierze blokowe X^{R_1} oraz X^{R_2} i przyjmijmy, że pierwsza z nich obejmuje n_1 jej początkowych wierszy macierzy, a druga n_2 wierszy pozostałych.

Dla każdego elementu $\rho^2 \in P^2$ statystycznym estymatorem m-wymiarowego wektora różnic $\delta(\rho^2)$ będzie zmienna losowa

$$\hat{X}(\rho^2) = \frac{1}{n_1}\sum_{s \in R_1} X_s - \frac{1}{n_2}\sum_{s \in R_2} X_s . \tag{4.45}$$

Oznaczmy przez $\hat{X}_1, \hat{X}_2, ..., \hat{X}_m$ jednowymiarowe zmienne losowe, będące składowymi wektora $\hat{X}(\rho^2)$, gdzie

$$\hat{X}_j = \frac{1}{n_1}\sum_{s \in R_1} X_{sj} - \frac{1}{n_2}\sum_{s \in R_2} X_{sj} , \tag{4.46}$$

Średnie, na podstawie których określa się $\hat{X}_j$, są estymatorami nieznanych wartości oczekiwanych Θ'_j i Θ_j $(j = 1,2,...,m)$.

Załóżmy, że rozpatrywane zmienne losowe $\hat{X}_j$ są normalne. Oznaczmy ich wariancje przez σ_j^2. Wtedy zmienne losowe $\frac{1}{n_1}\sum_{s \in R_1} X_{sj}$ i $\frac{1}{n_2}\sum_{s \in R_2} X_{sj}$ będą miały rozkłady normalne o wartościach średnich Θ'_j oraz Θ_j i wariancjach $\frac{\sigma^2}{n_1}$ oraz $\frac{\sigma^2}{n_2}$. Jest oczywiste, że zmienne losowe określone wzorem (4.46) – jako sumy zmiennych losowych normalnych – też mają rozkłady normalne, ale o parametrach

$$E(\hat{X}) = \Theta'_j - \Theta_j, \tag{4.47}$$

$$\hat{D}^2(\hat{X}) = \sigma_j^2 \frac{n_1 + n_2}{n_1 n_2}. \tag{4.48}$$

Empiryczne wartości zmiennych losowych $\hat{X}_j$ oblicza się z formuły

$$\hat{x}_j = \frac{1}{n_1}\sum_{s \in R_1} x_{sj} - \frac{1}{n_2}\sum_{s \in R_2} x_{sj}. \tag{4.49}$$

Wykorzystując powyższe rozważania można dla każdego elementu $\rho^2 \in P^2$, określonego hipotezą zerową (4.44), zbudować funkcję wiarygodności $L(\rho^2)$ i wykorzystać ją do skonstruowania kryterium weryfikacji hipotezy H_0. Jest zrozumiałe, że funkcją tą będzie w tym przypadku gęstość wielowymiarowego rozkładu normalnego, czyli funkcja

$$L(\rho^2) = (2\pi)^{-\frac{m}{2}} \left(\prod_{j=1}^{m} D^2(\hat{X}_j) \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^{m} \frac{[x_j - (\Theta'_j - \Theta_j)]^2}{D^2(\hat{X}_j)} \right]. \tag{4.50}$$

Kryterium jednorodności powinno stanowić podstawę do przyjęcia lub odrzucenia hipotezy zerowej H_0. Dlatego dla sprawdzenia jednorodności w pełni uzasadnione jest zbudowanie funkcji wiarygodności odpowiadającej sytuacji, gdy hipoteza H_0 jest prawdziwa. Jeśli przyjąć te warunki, to różnice $\Theta'_j - \Theta_j$ są równe zeru i funkcja wiarygodności przyjmie postać

$$L(\rho^2) = (2\pi)^{-\frac{m}{2}} \left(\prod_{j=1}^{m} D^2(\hat{X}_j) \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^{m} \frac{\hat{x}_j^2}{D^2(\hat{X}_j)} \right]. \quad (4.51)$$

O zachowaniu się funkcji (4.51) decyduje przede wszystkim wykładnik potęgi. Suma występująca w tym wykładniku jest wartością zmiennej losowej

$$\sum_{j=1}^{m} \frac{(\overline{X}'_j - \overline{X})^2}{D^2(\hat{X}_j)}. \quad (4.52)$$

Licznik ułamka stojącego pod znakiem sumy (4.52) jest kwadratem wyrażenia (4.49), zaś mianownik wyraża się wzorem

$$D\hat{X}_j = \sigma_j^2 \frac{n_1 + n_2}{n_1 n_2}. \quad (4.53)$$

Wyrażenie (4.52) jest więc sumą kwadratów m zmiennych losowych postaci

$$\frac{(\overline{X}'_j - X_j)^2}{\sigma_j} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (4.54)$$

o rozkładzie normalnym i zerowych wartościach oczekiwanych (ponieważ przyjęliśmy prawdziwość hipotezy zerowej) oraz wariancji równej 1. Stanowi to podstawę do uznania, że w przypadku prawdziwości hipotezy zerowej suma (4.52) ma rozkład χ^2 (chi-kwadrat) o m stopniach swobody.

A zatem, jako kryterium weryfikacji hipotezy o jednorodności dwóch zbiorów danych można przyjąć funkcję

$$U(\rho^2) = \frac{n_1 n_2}{n_1 + n_2} \sum_{j=1}^{m} \frac{(\overline{x}'_j - \overline{x}_j)^2}{\sigma_{ej}^2}, \quad (4.55)$$

gdzie: σ_{ej}^2 – empiryczna wartość wariancji σ_j^2 określona wzorem

$$\sigma_{ej}^2 = \frac{1}{n_1 + n_2 - 1} \left[\sum_{s \in R_1} x_{sj}^2 + \sum_{s \in R_2} x_{sj}^2 - \frac{1}{n_1 + n_2} \left(\sum_{s \in R_1} x_{sj} + \sum_{s \in R_2} x_{sj} \right)^2 \right], \quad (4.56)$$

przy czym estymatory $\bar{x}'_j$ i $\bar{x}_j$ są określone za pomocą odpowiednich sum stojących po prawej stronie równości (4.49).

Jeśli więc do formuły (4.55) podstawić wartości σ^2_{ej}, $\bar{x}'_j$ i $\bar{x}_j$ określone za pomocą (4.56) i (4.49), to – po wykonaniu niezbędnych przekształceń – kryterium jednorodności dwóch zbiorów tego samego zbioru przyjmie ostatecznie taką postać

$$U(\rho^2) = \frac{n_1 + n_2 - 1}{(n_1 + n_2)n_1 n_2} \sum_{j=1}^{m} \frac{\left(n_2 \sum_{s \in R_1} x_{sj} - n_1 \sum_{s \in R_2} x_{sj} \right)^2}{\sum_{s \in S} x_{sj}^2 - \frac{1}{n_1 + n_2} \left(\sum_{s \in S} x_{sj} \right)^2}. \tag{4.57}$$

Hipoteza o jednorodności analizowanych zbiorów (dwóch podzbiorów tego samego zbioru danych) może być przyjęta, jeśli dla każdego $\rho^2 \in P^2$

$$U(\rho^2) \leq \chi^2_{\alpha,m}, \tag{4.58}$$

gdzie α jest poziomem istotności, m – liczbą stopni swobody, $\chi^2_{\alpha,m}$ – wartość statystyki chi-kwadrat odczytana z tablic jej rozkład (dla danych α i m).

Jeśli nierówność (4.58) nie jest spełniona, to hipotezę o jednorodności należy odrzucić.

4.3.4. Warunki jednorodności – przypadek wielu zbiorów wielowymiarowych

Kryterium (4.57) jest dostosowane do sprawdzania jednorodności dwóch zbiorów X^{R_1} i X^{R_2} (w szczególności mogą to być podzbiory jednego i tego samego zbioru). Nie nadaje się jednak do testowania jednorodności więcj niż dwóch zbiorów (w tym kilku podzbiorów jednego zbioru) równocześnie. Co więcej, przy wyprowadzaniu tego kryterium przyjęliśmy, że testowanie jednorodności odbywa się na zbiorze danych otrzymanych bezpośrednio z obserwacji, tzn. nie poddanych żadnej wstępnej obróbce. W rzeczywistości częste są sytuacje, że oryginalny zbiór danych zostaje wstępnie przetworzony (np. jego elementy zostają uszeregowane w kolejności ich bezwzględnej wartości liczbowej lub w kolejności przypisanych im rang itp.). Kryterium (4.57) nie obejmuje talkich sytuacji. Powstaje więc potrzeba uogólnienia go w taki sposób, by nadawało się do testowania jednorodności wielu zbiorów równocześnie, nawet wtedy, gdy zostaną one poddane takim operacjom, jak szeregowanie, rangowanie itp. Zauważmy, że aby otrzymać takie uogólnienie wystarczy przyjąć, że:

1. Obserwowane wartości poszczególnych zmiennych losowych reprezentujących rozpatrywane cechy są uszeregowane w kolejności wzrastania (niemalenia) wartości jednej z nich, na przykład tej, która w danym badaniu została uznana za cechę dominującą.
2. W żadnej fazie procesu badania jednorodności struktura przetworzonego w ten sposób początkowego zbioru danych nie zostanie zmieniona (tzn. uszeregowanie zbioru realizacji m-wymiarowego wektora cech, w kolejności odpowiadającej wzrastaniu lub niemaleniu wartości współrzędnej odpowiadającej cesze dominującej, pozostaje niezmienne aż do zakończenia procesu badania jednorodności).

Zauważmy, że przyjęcie tych warunków nie pociąga za sobą utraty informacji zawartej w zbiorze danych, ani też nie komplikuje obliczeń związanych z jego obróbką statystyczną.

Zauważmy też, że jeżeli zbiór S zawiera n elementów, to zbiór wszystkich jego liniowo uporządkowanych podziałów zawiera $n-1$ elementów. Zbiór ten oznaczymy symbolem P_*^2, a jego element symbolem ρ_*^2.

Przyjmijmy, że X^s jest statystycznym modelem rozpatrywanego zbioru liniowo uporządkowanych obiektów. Ponieważ o podziale zbioru na części jednorodne można mówić tylko wtedy, gdy jest on niejednorodny, więc przede wszystkim trzeba rozstrzygnąć, czy zbiór X^s jest statystycznie jednorodny, czy też nie. Zakładamy, że zbiór ten jest zapisany w postaci macierzy (4.35). Wiersze tej macierzy są ponumerowane w kolejności do pierwszego do ostatniego za pomocą liczb naturalnych $i = 1,2,...,n$.

Oznaczmy podzbiór elementów zbioru S o numerach od 1 do l przez R_l, a jego dopełnienie do zbioru S (tj. zbiór S od $l+1$ do n) przez $\overline{R_l}$. Jest oczywiste, że dla $l = 1$ zbiorem X^{R_1} jest pierwszy wiersz macierzy (4.27), zaś pozostałe $n-1$ wierszy tworzy zbiór $X^{\overline{R}_1}$. Dla $l = 2$ zbiór X^{R_2} obejmuje dwa pierwsze wiersze macierzy (4.35), a pozostałe $n-2$ wiersze tworzą zbiór $X^{\hat{R}_2}$ itd.

Jeżeli są spełnione warunki 1 oraz 2 i jeżeli przyjmiemy, że rozpatrywane zmienne losowe są normalne, to hipoteza o jednorodności uporządkowanego zbioru danych statystycznych może być – przez analogię do (4.43) – sformułowana jako układ następujących $n-1$ hipotez (wymagających równoczesnego sprawdzenia)

$$H_0 : \delta(\rho_*^2) = \underbrace{(0,0,...0)}_{m} \quad \text{dla wszystkich} \quad \rho_*^2 \in P_*^2, \tag{4.59}$$

przy alternatywie

$$H_0 : \delta(\rho_*^2) \neq \underbrace{(0,0,...0)}_{m} \text{ dla co najmniej jednego } \rho_*^2 \in P_*^2 . \qquad (4.60)$$

W wyrażeniach (4.59) i (4.60) element ρ_*^2 jest określony na zbiorze podziałów P_*^2 zbioru X^s na wzajemnie dopełniające się podzbiory R_l i $\overline{R}_l$. Dlatego też teraz, dla obliczenia wartości $\delta(\rho_*^2)$ trzeba posłużyć się formułą (otrzymuje się ją analogicznie do (4.33), ale przyjmując, że $X^s = R_l \cup \overline{R}_l$)

$$\delta(\rho_*^2) = \frac{1}{l} \sum_{s \in R_l} \Theta_s - \frac{1}{n-l} \sum_{s \in \overline{R}_l} \Theta_s . \qquad (4.61)$$

W tym przypadku statystyka $U(\rho^2)$, którą posłużymy się do zweryfikowania hipotezy (4.59), przyjmuje postać (por. (4.57)

$$U(\rho^2) = \frac{n-1}{n(n-l)l} \sum_{j=1}^{m} \frac{\left((n-l)\sum_{s=1}^{l} x_{sj} - l \sum_{s=l+1}^{n} x_{sj} \right)^2}{\sum_{s=1}^{n} x_{sj}^2 - \frac{1}{n}\left(\sum_{s=1}^{n} x_{sj} \right)^2} . \qquad (4.62)$$

Jest ona określona na zbiorze podziałów ρ_*^2 i – w przypadku prawdziwości hipotezy zerowej - przedstawia wartość zmiennej losowej o rozkładzie chi-kwadrat o m stopniach swobody.

Wyniki obliczeń wedle formuły (4.62) uważa się za nieprzeczące możliwości przyjęciu hipotezy o jednorodności, jeśli jest spełniona nierówność

$$U(\rho_*^2) \leq \chi_{\alpha,m}^2 \text{ dla wszystkich } \rho_*^2 \in P_*^2 . \qquad (4.63)$$

gdzie α – poziom istotności, m – liczba stopni swobody, $\chi_{\alpha,m}^2$ – wartość statystyki chi-kwadrat odczytana z tablic jej rozkład (dla danych α i m).

Jeśli dla co najmniej jednego podziału $\rho_*^2 \in P_*^2$ ma miejsce $U(\rho_*^2) > \chi_{\alpha,m}^2$, to rozpatrywany zbiór – wobec odrzucenia hipotezy zerowej – nie może być uznany za jednorodny. W tym przypadku przyjmuje się więc hipotezę alternatywną (4.60) i należy rozwiązać zadaie podziału niejednorodnej zbiorowości statystycznej na grupy jednorodne.

Podsumowując, procedura weryfikacji hipotezy o jednorodności m-wymiarowej zbiorowości statystycznej wymaga wykonania następujących czynności:

1. Ustalenia, którą spośród m rozpatrywanych cech przyjmuje się za dominującą i uszeregowania początkowego zbioru danych w kolejności zgodnej z porządkiem wzrastania (niemalenia) wartości tej cechy. Uszeregowanie to jest utrzymane w toku całego badania.
2. Obliczenia sumy liczbowych wartości każdej cechy oraz sumy ich kwadratów (sumowanie po liczbie elementów rozpatrywanej zbiorowości).
3. Użycia tych sum do obliczenia wartości mianownika ułamka stojącego pod znakiem sumy w wyrażeniu (4.62). Wartość mianownika nie zmieni się w toku dalszych obliczeń.
4. Obliczenia wedle formuły (4.62) wartości statystyki $U(\rho^2)$ – kolejno dla podziałów macierzy na zbiory obejmujące l początkowych i $n-l$ pozostałych wierszy – i porównywania tych wartości na każdym kroku tych z odczytaną z tablic rozkładu chi-kwadrat wartością krytyczną $\chi^2_{\alpha,m}$ (formuła (4.63)). Jeśli wszystkie obliczone wartości $U(\rho^2)$ są mniejsze lub równe wartości krytycznej $\chi^2_{\alpha,m}$, to przyjmuje się, że rozpatrywana zbiorowość jest jednorodna.

4.3.5. Procedura dekompozycji m-wymiarowej niejednorodnej zbiorowości na grupy jednorodne

Jeżeli zostanie stwierdzone, że rozpatrywana m-wymiarowa zbiorowość studentów jest statystyczna niejednorodna, to – aby móc poddać ją jakiejkolwiek obróbce statystycznej – trzeba ją najpierw podzielić na podzbiory (zbiory, części, grupy) jednorodne. Można przyjąć, że podstawę takiego podziału powinnien stanowić ten wariant podziału teoretycznego, któremu odpowiada największe zróżnicowanie w zbiorze rozpatrywanych cech. Jeśli zaakceptować ten pogląd, to dla podzielenia całej zbiorowości na grupy statystycznie jednorodne wygodniej jest przyjąć jako hipotezę zerową postulat o niejednorodności. Hipotezą alternatywną będzie wtedy postulat, że rozważana zbiorowość jest jednorodna.

Warunkiem zaakceptowania hipotezy o niejednorodności zbiorowości jest spełnienie nierówności

$$\delta(\rho^2) \neq \underbrace{(0,0,\ldots 0)}_{m}. \qquad (4.64)$$

Zachodzi ona wtedy, gdy każdej parze podzbiorów $X^R, X^{\overline{R}}$ badanej zbiorowości odpowiada para wektorów wartości średnich $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_m)$ i $\Theta' = (\Theta_1', \Theta_2', \ldots, \Theta_m')$, istotnie różniących się od siebie. Funkcja wiarygodności dla wielowymiarowych zmiennych losowych $\overline{X} = \frac{1}{l}\sum_{s \in R} X_s$ i $\overline{X}' = \frac{1}{n-l}\sum_{s \in \overline{R}} X_s$, utworzonych z jednowymiarowych zmiennych losowych normalnych $\overline{X}_j$ i $\overline{X}'$ o wartościach średnich Θ_j i Θ_j' oraz wariancjach $\frac{\sigma_j^2}{l}$ i $\frac{\sigma_j^2}{n-l}$, jest określona wzorem

$$L(\rho^2) = \left[\frac{l(n-l)}{2\pi}\right]^{\frac{m}{2}} \prod_{j=1}^{m} \frac{1}{\sigma_j} \exp\left[-\frac{l(n-l)}{2n}\sum_{j=1}^{m} \frac{(x_j - \Theta_j)^2 + (x_j' - \Theta_j')^2}{\sigma_j^2}\right]. \quad (4.65)$$

Z istoty metody największej wiarygodności wynika, że w przypadku prawdziwości hipotezy o niejednorodności funkcja $L(\rho^2)$ musi osiągać maksimum. Z postaci tej funkcji wnioskujemy, że będzie to miało miejsce wtedy, gdy wykładnik występującej w niej funkcji eksponencjalnej przyjmie wartość minimalną. Innymi słowy, w przypadku, gdy $\Theta_j \neq \Theta_j'$ $(j = 1,2,\ldots,n)$, wyrażenie

$$-\frac{l(n-l)}{n}\sum_{j=1}^{m} \frac{(x_j - \Theta_j)^2 + (x_j' - \Theta_j')^2}{\sigma_j^2}. \quad (4.66)$$

musi osiągać minimum względem x_j i x_j'. Aby znaleźć to minimum rozpatrzmy najpierw licznik ułamka stojącego w (4.66) pod znakiem sumy, tj. wyrażenie

$$(x_j - \Theta_j)^2 + (x_j' - \Theta_j')^2. \quad (4.67)$$

Przypuśćmy, że chcemy znaleźć jego minimum względem x_j i x_j'. Ponieważ rownoważną postacią tego wyrażenia jest

$$\sum_{s\in R}(x_{sj}-\Theta_j)^2+\sum_{s\in\overline{R}}(x_{sj}-\Theta'_j)^2=\sum_{s\in R}(x_{sj}-x_j)^2+n_R(x_j-\Theta_j)^2+ \\ +\sum_{s\in\overline{R}}(x_{sj}-x'_j)^2+n_{\overline{R}}(x'_j-\Theta'_j)^2 \tag{4.68}$$

gdzie n_R - liczba elementów w zbiorze R, $\overline{n}_R$ - liczba elementów w zbiorze $\overline{R}$, więc minimalizację formuły (4.67) można zastąpić minimalizacją formuły (4.68). Zapiszmy formułę (4.68) w postaci

$$\sum_{s\in R}(x_{sj}-x_j)^2+\sum_{s\in\overline{R}}(x_{sj}-x'_j)^2 \tag{4.69}$$

Ponieważ

$$\sum_{s\in R}(x_{sj}-\overline{x})^2+\sum_{s\in\overline{R}}(x_{sj}-\overline{x})^2=\sum_{s\in R}(x_{sj}-x_j)^2+n_R(\overline{x}-x_j)^2+ \\ +\sum_{s\in\overline{R}}(x_{sj}-x'_j)^2+n_{\overline{R}}(\overline{x}-x'_j)^2, \tag{4.70}$$

więc wyrażenie (4.69) osiąga wartość minimalną wtedy, gdy wyrażenie

$$n_R(\overline{x}-x_j)^2+n_{\overline{R}}(\overline{x}-x'_j)^2 \tag{4.71}$$

przyjmuje wartość maksymalną. Zapiszmy formułę (4.71) w takiej postaci

$$\frac{n_R}{n}(\overline{x}-x_j)^2+\frac{n_{\overline{R}}}{n}(\overline{x}-x'_j)^2 \tag{4.72}$$

i wprowadźmy następującą notację: $x_j=A$, $x'_j=B$, $\frac{n_R}{n}=p$, $\frac{n_{\overline{R}}}{n}=q$, $\overline{x}=Ap+Bq=\gamma$. Ponieważ

$$p(A-\gamma)^2+q(B-\gamma)^2=(A-B)^2pq, \tag{4.73}$$

więc maksymalizacja wyrażenia (4.72) sprowadza się do maksymalizacji wyrażenia $(A-B)^2pq$, lub, co na jedno wychodzi, wyrażenia

$$\frac{n_R n_{\overline{R}}(x_j-x'_j)^2}{n^2}. \tag{4.74}$$

Rozpatrzmy wyraźnie

$$\frac{n_R n_{\bar{R}} (x_j - x'_j)^2}{\sigma_j^2 n}. \tag{4.75}$$

Zwróćmy uwagę na jego formalne podobieństwo do wyrażenia (4.54). Zauważmy jednak, że wyrażenie (4.75) określa maksymalną wartość wyrażenia (4.65). Jeżeli w formule (4.75) zastąpimy wielkości x_j i x'_j wyrażeniami (4.49), a wariancję σ_j^2 formułą (4.48), to – po wykonaniu prostych przekształceń – ostatecznie otrzymamy taką postać wyrażenia na obliczanie maksymalnej wartości kryterium podziału niejednorodnej zbiorowości na grupy jednorodne

$$U_{\max}(\rho^2) = \frac{n-1}{n(n-l)l} \sum_{j=1}^{m} \frac{\left((n-l) \sum_{s=1}^{l} x_{sj} - l \sum_{s=l+1}^{n} x_{sj} \right)^2}{\sum_{s=1}^{n} x_{sj}^2 - \frac{1}{n} \left(\sum_{s=1}^{n} x_{sj} \right)^2}. \tag{4.76}$$

Wyrażenie to mówi, że najbardziej wiarygodny jest ten podział niejednorodnej m-wymiarowej zbiorowości statystycznej na dwie części, któremu odpowiada największa wartość statystyki $U(\rho^2)$. A zatem, aby podzielić tę zbiorowość na zbiory jednorodne, trzeba posłużyć się statystyką (4.62) i skorzystać z wyniku (4.76). Procedura podziału będzie realizowana w następujący sposób:

1. Dla danej zbiorowości statystycznej (uporządkowanej według cechy przyjętej za dominującą) oblicza się wartości statystyki $U(\rho^2)$ i przedstawia się je w postaci odpowiedniej tablicy. W tablicy tej odnajduje się maksymalną wartość statystyki $U(\rho^2)$, nie spełniającą warunku (4.63). Wiersz, w którym ona występuje wyznacza podział zbiorowości na dwie części (Tablica 4.2).

2. Sprawdza się jednorodność każdej z tych części. Jeśli obie one są jednorodne, to podział uważa się za skończony. Jeśli jednak dla którejkolwiek z tych części hipoteza o jednorodności zostanie odrzucona, to tę część początkowej zbiorowości trzeba podzielić na dwie części (kierując się maksymalną wartością kryterium (4.76), a następnie sprawdzić jednorodność każdego z otrzymanych zbiorów. Postępowanie to prowadzi się dopóty, dopóki nie przestaną występować zbiory niejednorodne, tzn. dopóki nie zostaną określone wszystkie wartości statystyki $U(\rho^2)$, wyznaczające progi podziału zbiorowości na grupy jednorodne.

Otrzymane za pomocą tej procedury statystycznie jednorodne zbiory mogą wyraźnie różnić się między sobą, ale może też zdarzyć się, że różnice między niektórymi z nich nie są istotne. Powstaje więc potrzeba sprawdzenia stopnia istotności różnicy między każdymi dwoma sąsiednimi zbiorami jednorodnymi, otrzymanymi w toku podziału. W przypadku, gdy różnica ta między dwoma sąsiednimi zbiorami okaże się nieistotna, należy zbiory te połączyć ze sobą.

Statystyczną istotność różnic między parami sąsiednich zbiorów jednorodnych można sprawdzić porównując odpowiadające im średnie wielowymiarowe. Jeśli okaże się, że numerycznie są one sobie równe, to można przyjąć, że różnica między zbiorami jest nieistotna i połączyć te zbiory w jeden. Jeśli jednak okaże się, że wartości średnich wyraźnie różnią się między sobą, to trzeba zaakceptować odrębność zbiorów i nie wolno ich ze sobą łączyć.

Aby skonstruować kryterium pozwalające rozstrzygnąć, czy sąsiadujące ze sobą zbiory jednorodne istotnie różnią się między sobą przyjmijmy, że rozpatrywana m-wymiarowa zbiorowość studentów została podzielona na q grup jednorodnych. Niech Θ_k $(k=1,2,...,q)$ będzie średnią m-wymiarową w k-tej grupie. Jako hipotezę zerową H_0 przyjmijmy stwierdzenie, że różnica między grupami X^{S_k} i $X^{S_{k+1}}$ jest statystycznie nieistotna. Formalnie odpowiada to zapisowi

$$H_0 : \Theta_k - \Theta_{k+1} = \underbrace{(0,0,...,0)}_{m} . \tag{4.77}$$

Hipotezą alternatywną jest

$$H_1 : \Theta_k - \Theta_{k+1} \neq \underbrace{(0,0,...,0)}_{m} . \tag{4.78}$$

Przyjęcie hipotezy zerowej wymaga uznania, że różnica między grupami X^{S_k} i $X^{S_{k+1}}$ jest statystycznie nieistotna i wobec tego grupy te łączy się w jedną $(X^{S_k} \cup X^{S_{k+1}})$. Odrzucenie hipotezy (4.77) sprzyja przyjęciu jej alternatywy (4.78), a więc uznania, że grupy X^{S_k} i $X^{S_{k+1}}$ różnią się od siebie w sposób statystycznie istotny. Utrzymuje się więc odrębność tych grup i przechodzi do weryfikowania hipotezy

$$H_0 : \Theta_{k+1} - \Theta_{k+2} = \underbrace{(0,0,...,0)}_{m} , \tag{4.79}$$

tj. do porównania m-wymiarowych średnich następnej pary sąsiednich grup przy alternatywie

$$H_0 : \Theta_{k+1} - \Theta_{k+2} \neq \underbrace{(0,0,...,0)}_{m} . \tag{4.80}$$

Postępowanie to kontynuuje się aż do wyczerpania wszystkich par sąsiednich grup.

Aby skonstruować kryterium (statystykę) dla sprawdzania hipotez (4.77)-(4.80) o statystycznej istotności lub nieistotności różnic między jednorodnymi grupami, otrzymanymi w wyniku podziału rozpatrywanej zbiorowości, trzeba przeprowadzić rozumowanie analogiczne do tego, które prowadziło do otrzymania kryterium (4.62). Jeżeli przyjmie się, że $S = S_1 \cup S_2 \cup \ldots \cup S_q$, $S_r \cap S_s = \emptyset \; (r, s = 1,2,\ldots,q; r \neq s)$, to otrzyma się następującą formułę

$$U(S_k, S_{k+1}) = \frac{n_k + n_{k+1} - 1}{n_k n_{k+1}(n_k + n_{k+1})} \sum_{j=1}^{m} \frac{\left(n_{k+1} \sum_{s \in S_k} x_{sj} - n_k \sum_{s \in S_{k+1}} x_{sj} \right)^2}{\sum_{s \in S_k \cup S_{k+1}} x_{sj}^2 - \frac{1}{n_k + n_{k+1}} \left(\sum_{s \in S_k \cup S_{k+1}} x_{sj} \right)^2}, \tag{4.81}$$

gdzie $q = 1,2,\ldots,q-1$.

Po obliczeniu dla każdej pary S_k, S_{k+1} $(k = 1,2,\ldots,q-1)$ wartości statystyki (4.81) porównuje się ją z wartością $\chi^2_{\alpha,m}$ odczytaną tablic rozkładu chi-kwadrat. Jeśli okaże się, że $U(S_k, S_{k+1}) < \chi^2_{\alpha,m}$, to przyjmuje się, że różnica między zbiorami S_k, S_{k+1} nie jest istotna. Grupy te łączy się ze sobą i otrzymany zbiór porównuje się z grupą S_{k+2} itd. Jeśli jednak $U(S_k, S_{k+1}) > \chi^2_{\alpha,m}$ to odrębność zbiorów S_k, S_{k+1} zostaje zachowana i przechodzi się do weryfikacji istotności różnicy między grupami S_{k+1}, S_{k+2}. Jeśli stwierdzi się, że różnica ta jest niestotna, to grupy S_{k+1}, S_{k+2} scala się w jedną i przechodzi się do testowania istotności różnicy między nią, a grupą S_{k+3} itd.

4.3.6. Analiza jednowymiarowa jako szczególny przypadek wielowymiarowej

Spójrzmy teraz na analizę statystycznej jednorodności zbioru danych jednowymiarowych (każdy student lub nauczyciel jest opisany za pomocą jednej cechy skalarnej) jako na szczególny przypadek analizy zbioru m-wymiarowego (każdy student lub nauczyciel jest opisany za pomocą m-wymiarowego wektora cech). Główne formuły matematyczne dla przypadku skalarnego otrzymuje się z formuł (4.57), (4.62) i (4.76), odpowiadających przypadkowi m-wymiarowemu, przez położenie w tych ostatnich $m = 1$. W rezultacie otrzymuje się, że kryterium statystycznej jednorodności pary zbiorów jednowymiarowych S_k, S_{k+1} $(k = 1,2,\dots,q-1)$ o licznościach, odpowiednio, n_k, n_{k+1} jest – zgodnie z formułą (4.57) – statystyka

$$U(\rho^2) = \frac{n_1 + n_2 - 1}{n_1 n_2 (n_1 + n_2)} \frac{\left(n_2 \sum_{s \in R_1} x_s - n_1 \sum_{s \in R_2} x_s \right)^2}{\sum_{s \in S} x_s^2 - \frac{1}{n_1 + n_2} \left(\sum_{s \in S} x_s \right)^2}. \tag{4.82}$$

Statystyce (4.62), służącej do sprawdzania jednorodności na zbiorze podziałów zbiorowości wielowymiarowej, odpowiada w przypadku jednowymiarowym formuła

$$U(\rho^2) = \frac{n-l}{n(n-l)l} \frac{\left((n-l) \sum_{s \in R_1} x_s - l \sum_{s \in R_2} x_s \right)^2}{\sum_{s \in S} x_s^2 - \frac{1}{n} \left(\sum_{s \in S} x_s \right)^2}, \tag{4.83}$$

gdzie $l = 1,2,\dots,n-1$.

I wreszcie, wyrażeniu (4.76) na maksymalną wartość statystyki $U(\rho^2)$, wyznaczającą – zgodnie z kryterium największej wiarygodności – miejsce podziału m-wymiarowej zbiorowości na dwie części, odpowiada w przypadku jednowymiarowym formuła

$$U_{\max}(\rho^2) = \frac{n-l}{n(n-l)l} \frac{\left((n-l) \sum_{s \in R_1} x_s - l \sum_{s \in R_2} x_s \right)^2}{\sum_{s \in S} x_s^2 - \frac{1}{n} \left(\sum_{s \in S} x_s \right)^2}, \quad (l = 1,2,\dots,n-1). \tag{4.84}$$

Jest oczywiste, że procedura podziału w przypadku jednowymiarowym jest w zasadzie powtórzeniem procedury odpowiadającej przypadkowi wielowymiarowej (por. formuły (4.59), (4.62) i (4.76)).

Zauważmy jeszcze, że kryterium (4.81), którego wartość liczbowa w przypadku m-wymiarowym decyduje, czy sąsiadujące ze sobą grupy jednorodne można połączyć w jedną grupę czy też nie, w przypadku zbiorowości jednowymiarowej przyjmuje postać

$$U(S_k, S_{k+1}) = \frac{n_k + n_{k+1} - 1}{n_k n_{k+1}(n_k + n_{k+1})} \sum_{j=1}^{m} \frac{\left(n_{k+1} \sum_{s \in S_k} x_s - n_k \sum_{s \in S_{k+1}} x_s \right)^2}{\sum_{s \in S_k \cup S_{k+1}} x_s^2 - \frac{1}{n_k + n_{k+1}} \left(\sum_{s \in S_k \cup S_{k+1}} x_s \right)^2}, \tag{4.85}$$

gdzie $k = 1,2,...,q-1$.

5. Przykład zastosowania

Załóżmy, że na pierwszy rok studiów technicznych przyjęto 120 osób. Podzielono je administracyjnie na 6 dwudziestoosobowych grup dydaktycznych Dla celów analizy, aby nie operować nazwiskami, studentom przyporządkowano losowo numery porządkowe od 1 do 120. Podział na grupy jest przedstawiony w tablicy 5.1.

Tablica 5.1
Podział administracyjny zbiorowości studentów na grupy dydaktyczne

Lp.	Grupa					
	I	II	III	IV	V	VI
1	1	3	4	7	9	13
2	2	5	8	12	17	21
3	6	10	11	16	18	35
4	14	15	20	27	29	38
5	23	19	26	28	37	43
6	24	22	32	34	41	44
7	25	30	40	39	48	45
8	33	31	42	47	54	60
9	36	57	49	52	56	61
10	46	58	50	53	64	68
11	55	62	51	59	65	71
12	67	69	79	63	78	72
13	73	75	80	66	84	77
14	81	76	82	70	85	95
15	92	86	83	74	88	96
16	100	91	87	94	89	105
17	109	97	90	102	98	106

18	111	101	93	103	99	113
19	118	104	108	112	115	114
20	119	107	110	120	116	117

W toku zajęć okazało się, że zróżnicowanie merytorycznego poziomu studentów w każdej z grup jest zbyt duże: zarówno studenci dobrze przygotowani do studiów, jak i ci o słabym przygotowaniu zdobywają nową wiedzę w tempie dużo wolniejszym od przewidywanego w programie studiów, ponieważ poziom i tempo prowadzenia zajęć muszą być dostosowane do wiedzy i percepcyjnych możliwości przeciętnego członka grupy. Doświadczenie z poprzednich lat uczyło, że taka sytuacja rodzi poważne problemy: studenci występują o przeniesienie do innych grup, wyniki kolokwiów nie odpowiadają faktycznemu stanowi wiedzy studentów, zbyt wielu studentów ma kłopoty z zaliczeniem ćwiczeń i zdaniem egzaminów, dochodzi do konfliktów między studentami i wykładowcami itp. Wielu studentów słabszych ubiega się o egzaminy komisyjne, zawiesza studia na pewien czas lub rezygnuje z ich kontynuowania. Co gorsza, zaobserwowano też, że wcale nie mała liczba studentów dobrych przenosi się na inne uczelnie, w których poziom studiów bardziej odpowiada ich aspiracjom. Wszystko to wpływa na kondycję i prestiż uczelni, bowiem informacje o jakości studiów i warunkach studiowania rozchodzą się wśród młodzieży błyskawicznie.

Przyjmijmy, że w ciągu kilku ostatnich lat zaobserwowano gwałtowny spadek zainteresowania studiami, a przyczyn tego stanu rzeczy upatruje się przede wszystkim w złej organizacji i złej jakości procesu edukacyjnego. W wyniku przeprowadzonych analiz postanowiono radykalnie zmienić organizację tego procesu zwłaszcza na pierwszych dwóch latach studiów. Przyjęto, że zdecydowana większość młodzieży podejmującej studia nie chce jedynie przebrnąć przez nie i uzyskać dyplom ich ukończenia, lecz pragnie zdobyć wiedzę i zawód. Przyjęto też, że zadaniem uczelni jest stworzenie warunków, by spełnić te oczekiwania. Stwierdzono, że wymaga to, przede wszystkim, odejścia od tradycyjnie stosowanego administracyjnego trybu formowania grup studenckich. Postanowiono, że w pierwszym semestrze studiów zajęcia z matematyki oraz fizyki będą poświęcone solidnemu powtórzeniu i uzupełnieniu materiału licealnego z tych przedmiotów (część studentów miała w liceach podstawowy zakres matematyki, część uczyła się wedle zakresu rozszerzonego), przy czym zajęcia miały być tak ukierunkowane, by przygotowywały młodzież do studiowania przedmiotów wymagających odpowiedniego stopnia znajomości matematyki i fizyki. Zajęcia kończyły się sprawdzianem kompetencji, któremu musieli poddać się wszyscy studenci. Założymy, przykładowo, że składał się on z trzech części. Pierwsza dotyczyła matematyki, druga – fizyki, trzecia miała charakter ogólny. Części pierwsza i druga miały taką samą strukturę i zawierały po 30 zadań zamkniętych oraz 5 zadań otwartych. Część trzecia zawierała 30 zadań zamkniętych, których celem było sprawdzenie ogólnego poziomu erudycji (poza matematyką i fizyką) studenta. Rozwiązywanie zadań zamkniętych polegało na zaznaczeniu tej spośród czterech wersji odpowiedzi, którą rozwiązujący uzna za prawidłową, przy czym tylko jedna z nich była prawdziwa.

Zadania zamknięte były oceniane w systemie zerojedynkowym (1 punkt za odpowiedź prawidłową, 0 punktów za nieprawidłową lub zaznaczenie więcej niż jednej odpowiedzi, nawet gdyby wśród nich została zaznaczona odpowiedź poprawna). Tak więc w każdej części sprawdzianu za rozwiązanie zadań zamkniętych student mógł otrzymać od 0 do 30 punktów. Za rozwiązanie zadań otwartych (części matematyczna i fizyczna) można było otrzymać od 0 do 30 punktów, przy czym w każdym z zadań były punktowane logika rozumowania oraz operacje świadczące o opanowaniu przez studenta umiejętności stosowania posiadanej wiedzy do rozwiązywania zadań praktycznych. Za część matematyczną student mógł więc otrzymać łącznie od 0 do 120 punktów. Taką samą liczbę punktów mógł otrzymać za rozwiązanie części fizycznej.

W wyniku sprawdzianu każdemu studentowi została przyporządkowana uporządkowana trójka liczb (x_1, x_2, x_3), gdzie x_1 - łączna liczba punktów za część matematyczną, x_2 - łączna liczba punktów za część fizyczną, x_3 - liczba punktów za część ogólną. Wyniki te posłużyły do przeanalizowania stopnia merytorycznej jednorodności zbiorowości studentów z punktu widzenia tych trzech wskaźników równocześnie, a następnie do dekompozycji tej zbiorowości (reprezentowanej przez trójwymiarowy zbiór danych) na grupy statystycznie jednorodne.

Procedura sprawdzania jednorodności zbiorowości studentów i podziału na grupy jednorodne była realizowana wedle następującego schematu, odpowiadającego przedstawionej idei badania jednorodności:

Krok 1

Zapisanie wyników sprawdzianu w postaci początkowej macierzy danych (wzór (4.35) liczącej 120 wierszy (liczba studentów) i 3 kolumny (liczba cech charakteryzujących każdego studenta).

Krok 2

Wybór cechy dominującej – przyjęto, że jest nią x_1, tzn. łączna liczba punktów otrzymanych za matematyczną część sprawdzianu.

Krok 3

Skonstruowanie zmodyfikowanej macierzy danych – modyfikacja polega na ustawieniu wierszy macierzy początkowej w kolejności odpowiadającej porządkowi niemalenia wartości cechy x_1. Dalsze obliczenia są prowadzone w odniesieniu do tej macierzy.

Krok 4

Obliczenie dla zmodyfikowanej macierzy danych sumy elementów oraz sumy kwadratów elementów stojących w poszczególnych kolumnach. Użycie tych sum do obliczenia wartości mianownika ułamka stojącego pod znakiem sumy w wyrażeniu (4.62). Wartość mianownika nie zmienia się w toku dalszych obliczeń.

Krok 5

Dzielenie zmodyfikowanej macierzy danych kolejno na macierze zawierające pierwszy wiersz i wszystkie wiersze pozostałe, dwa pierwsze wiersze i wszystkie wiersze pozostałe, trzy pierwsze wiersze i wszystkie wiersze pozostałe itd., aż do wyczerpania możliwości podziału i obliczenie – za każdym razem – wartości statystyki $U(\rho^2)$ wedle formuły (4.62).

Krok 6

Ustalenie wartości poziomu istotności α (przyjęto $\alpha = 0.05$) i porównanie otrzymanej wartości $U(\rho^2)$ z wartością krytyczną $\chi^2_{0.05,3} = 0.7815$, odczytaną z tablic rozkładu chi-kwadrat (formuła (4.63)). Ponieważ najmniejsza z wartości $U(\rho^2)$ okazała się większa od 0.7815, więc wnioskuje się, że rozpatrywana populacja studentów (trójwymiarowy zbiór danych) nie może być uważana za statystycznie jednorodną.

Procedurę dekompozycyjną zatrzymano, gdy zostały zidentyfikowane wszystkie jednorodne podzbiory rozpatrywanej zbiorowości studentów – otrzymano łącznie 10 takich podzbiorów. Zgodnie z ideą metody przystąpiono więc do sprawdzenia, czy w otrzymanym podziale różnice między sąsiadującymi ze sobą podzbiorami jednorodnymi są statystycznie istotne, czy też nie. W przypadku stwierdzenia nieistotności zbiory te łączono ze sobą. Realizacja tego postępowania przebiegała tak:

Krok 1

Rozpatruje się dwie pierwsze grupy jednorodne. Oblicza się dla nich wartość statystyki (4.81). Wartość tę porównuje się z wartością $\chi^2_{0.05,3} = 7.815$, odczytaną z tablic rozkładu chi-kwadrat. Jeżeli $U(S_1, S_2) > 7.815$, to przyjmuje się, że grupy S_1 i S_2 różnią się istotnie i ich odrębność zostaje zachowana (w przeciwnym razie należałoby połączyć je ze sobą).

Krok 2

Rozpatruje się grupę drugą i trzecią, następnie trzecią i czwartą itd., aż do wyczerpania możliwości dalszej analizy. Za każdym razem oblicza się wartość statystyki (4.81) i porównuje ją z wartością $\chi^2_{0.05,3} = 7.815$. Wynik porównania decyduje o utrzymaniu odrębności grup lub o ich połączeniu w jedną.

Tablica 5.2

Wyniki ostatecznego podziału zbiorowości studentów na grupy statystycznie jednorodne w sensie wskaźników x_1, x_2 i x_3

Lp.	Grupa				Lp.	Grupa			
	I	II	III	IV		I	II	III	IV
1	1	4	2	8	36	86	112		
2	3	11	6	13	37	87			
3	5	12	7	16	38	88			
4	10	15	9	19	39	90			
5	14	17	21	29	40	92			
6	18	20	26	39	41	94			
7	22	25	31	43	42	95			
8	23	27	33	71	43	99			
9	24	28	37	96	44	100			
10	30	32	38	101	45	103			
11	34	40	50	102	46	104			
12	35	41	97	110	47	105			
13	36	44	106		48	107			
14	42	48	111		49	108			
15	45	49			50	109			
16	46	51			51	113			
17	47	55			52	1114			
18	52	57			53	115			
19	53	59			54	116			
20	54	60			55	117			
21	56	61			56	118			
22	58	62			57	119			
23	65	63			58	120			
24	66	64							
25	67	69							
26	68	74							
27	70	79							
28	72	80							
29	73	81							
30	75	82							

31	76	84							
32	77	89							
33	78	91							
34	83	93							
35	85	98							

Wyniki otrzymane po zbadaniu, czy sąsiadujące ze sobą grupy jednorodne różnią się od siebie istotnie, czy też nie i – w przypadku stwierdzenia nieistotności - połączeniu ich ze sobą, są przedstawione w tablicy 5.2. Okazało się, że rozpatrywana zbiorowość studentów jest silnie niejednorodna i daje się ostatecznie podzielić na cztery istotnie różniące się między sobą grupy jednorodne o licznościach – odpowiednio – 58 (słaby wyniki sprawdzianu), 36 (dostateczny wynik sprawdzianu), 14 (dobry wynik sprawdzianu) 12 (bardzo dobry wynik sprawdzianu) osób. Składy osobowe grup są podane w tablicy 5.2.

Otrzymane wyniki mogą być podstawą nowego podziału zbiorowości studentów na grupy. I tak, z pierwszej grupy jednorodnej można utworzyć trzy grupy zajęciowe liczące – odpowiednio – 20, 20 i 18 osób. Drugą grupę jednorodną można podzielić na dwie grupy zajęciowe, liczące po 18 osób. Trzecia i czwarta grupa jednorodna powinny pozostać niezmienione. Łącznie powstałoby więc siedem grup zajęciowych. W drugim semestrze studenci słabi i średni powinni obowiązkowo uczestniczyć w dalszych zajęciach wyrównawczych z matematyki i fizyki (np. w wymiarze dwóch godzin zajęć co dwa tygodnie).

6. Statystyka T^2 Hotellinga a statystyka U

Problem skonstruowania statystycznego kryterium jednorodności dwóch wielowymiarowych zbiorów danych nie jest nowy. W przeszłości zajmowali się nim statystycy tej miary co Hotelling (1951), Anderson i Darling (1952), Fisher (1958), Anderson (1962), Kullback (1959), Lehmann (1959), Day (1969) i wielu innych. U podstaw prowadzonych przez nich badań leży następująca statystyka, wprowadzona przez Hotellinga:

$$T^2 = n(\bar{x} - \mu)' V^{-1} (\bar{x} - \mu), \qquad (6.1)$$

gdzie $\bar{x}$ jest wektorem wartości średnich, μ – wartością oczekiwaną, V – macierzą kowariancyjną zbioru danych zawierającego n elementów.

Szczególnie ważny jest wynik Andersona (1958), który badał rozkład tej statystyki i udowodnił, że – ze wszystkich kryteriów służących do sprawdzania prawdziwości hipotezy zerowej – największą moc ma kryterium oparte na statystyce T^2. Posługując się tym kryterium Anderson rozwiązał zadanie testowania jednorodności dwóch lub więcej zbiorów danych i wykazał, że w warunkach prawdziwości hipotezy zerowej kryterium to ma rozkład F. Ito (1959) udowodnił, że formuła asymptotyczna dla rozkładu statystyki T^2 daje się wyrazić za pomocą rozkładu chi-kwadrat. Wynik ten pozwala spojrzeć na zadanie budowy kryterium jednorodności dwóch odrębnych zbiorów danych z tego samego punktu widzenia, co przyjęty w rozdz. 4.

Założmy, że badana zbiorowość została podzielona na q grup i że rozpatrywane zmienne losowe są niezależne i normalne. Statystycznymi modelami tych grup są jednorodne zbiory X^{K_l} zmiennych losowych X_k, gdzie $k \in K_l$. Niech X^{K_q} oznacza klasę rozłącznych zbiorów X^{K_l}, powstałych przy podziale przestrzeni X^K na q grup. Przestrzeń tę będziemy uważać za jednorodną, jeśli dla wszystkich $k, k' \in K$ jest spełniony warunek

$$\Theta_k - \Theta_{k'} = (\underbrace{0,0,\ldots,0}_{m}) . \tag{6.2}$$

Z równości tej wynika, że jeśli dla wszystkich zbiorów X^{K_l} należących do X^{K_q} i dla wszystkich $k, k' \in K_l$ zachodzi warunek

$$\Theta_k - \Theta_{k'} = (\underbrace{0,0,\ldots,0}_{m}) , \tag{6.3}$$

to wymóg jednorodności klasy x^{K_q} jest równoważny wymogowi jednorodności przestrzeni X^K.

Oznaczmy wartość średnią zmiennej losowej X_k $(k \in K_l)$ przez Θ_l $(l = 1,2,\ldots,q)$, a wartość średnią zmiennej losowej X_k $(k \in K_t)$ przez Θ_t $(t = 1,2,\ldots,q)$. Jeżeli więc X^{K_q} jest jednorodną klasą zbiorów X^{K_l}, to jest spełniony następujący układ $\frac{q(q-1)}{2}$ równań:

$$\Theta_l - \Theta_t = (\underbrace{0,0,\ldots,0}_{m}), \qquad (l, t = 1,2,\ldots,q) . \tag{6.4}$$

Aby skonstruować kryterium jednorodności rozpatrzmy zmienną losową

$$X_{lt} = \frac{1}{n_l} \sum_{k \in K_l} X_k - \frac{1}{n_t} \sum_{k \in K_t} X_k , \tag{6.5}$$

(oczywiście $X_{lt} = (\hat{X}_{lt1}, \hat{X}_{lt2}, \ldots, \hat{X}_{ltm})$), gdzie n_l i n_t, są – odpowiednio – liczbami elementów zbiorów X^{K_l} i X^{K_t}, zaś

$$\hat{X}_{ltj}=\frac{1}{n_l}\sum_{k\in K_l}\hat{X}_{kj}-\frac{1}{n_t}\sum_{h\in K_t}\hat{X}_{kj}. \tag{6.6}$$

Zmienne losowe $\frac{1}{n_l}\sum_{h\in K_l}\hat{X}_{kj}$ i $\frac{1}{n_t}\sum_{h\in K_t}\hat{X}_{kj}$ są estymatorami wartości oczekiwanych Θ_l i Θ_t. Zgodnie z przyjętym wyżej założeniem zmienne te mają rozkłady normalne o parametrach rozkładu Θ_l i Θ_t oraz $\frac{\sigma_{kj}^2}{n_l}$ i $\frac{\sigma_{kj}^2}{n_t}$. Wobec tego, zmienne losowe $\hat{X}_{ltj}$ też mają rozkłady normalne, ale o wartościach oczekiwanych

$$E(\hat{X}_{ltj})=\Theta_l-\Theta_t \tag{6.7}$$

i wariancjach

$$D(\hat{X}_{ltj})=\sigma_{kj}^2\frac{n_l+n_t}{n_l n_t}. \tag{6.8}$$

Dla obliczenia empirycznych wartości zmiennych losowych $\hat{X}_{ltj}$ trzeba posłużyć się formułą

$$\hat{x}_{ltj}=\frac{1}{n_l}\sum_{k\in K_l}x_{kj}-\frac{1}{n_t}\sum_{h\in K_t}x_{kj}, \tag{6.9}$$

natomiast empiryczne oceny wariancji σ_{kj}^2 (oznaczymy je symbolem σ_{ekj}^2) oblicza się korzystając z wyrażenia

$$\sigma_{ekj}^2=\frac{1}{n_l+n_t-1}\left[\sum_{k\in K_l}x_{kj}^2+\sum_{k\in K_l}x_{kj}-\frac{1}{n_l+n_t}\left(\sum_{k\in K_l}x_{kj}+\sum_{k\in K_t}x_{kj}\right)^2\right]. \tag{6.10}$$

Jeśli skorzystamy z metody największej wiarygodności, to gęstość prawdopodobieństwa łącznego rozkładu zmiennych losowych $\hat{X}_{lt1},\bar{X}_{lt2},...,\hat{X}_{ltm}$ wyrazi się wzorem

$$L_{lt} = (2\pi)^{-\frac{m}{2}} \left(\prod_{j=1}^{m} D(\hat{X}_{ltj}) \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^{m} \frac{[x_{ltj} - (\Theta_{lj} - \Theta_{tj})]^2}{D(\hat{X}_{ltj})} \right]. \quad (6.11)$$

Jeśli hipoteza zerowa jest słuszna, to zgodnie z równością (5.4) różnica $\Theta_l - \Theta_t = 0$ i w tym przypadku funkcja wiarygodności przyjmie postać:

$$L_{lt} = (2\pi)^{-\frac{m}{2}} \left(\prod_{j=1}^{m} D(\hat{X}_{ltj}) \right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^{m} \frac{x_{ltj}^2}{D(\hat{X}_{ltj})} \right]. \quad (6.12)$$

Z metody największego prawdopodobieństwa wynika, że przyjęciu hipotezy o jednorodności najbardziej odpowiada ta para zbiorów X^{K_l} i X^{K_t}, dla której funkcja (6.12) osiąga wartość największą, lub, co na jedno wychodzi, gdy suma

$$\sum_{j=1}^{m} \frac{x_{ltj}^2}{D(\hat{X}_{ltj})}, \quad (6.13)$$

osiąga wartość najmniejszą. Zauważmy, że suma ta jest wykładnikiem potęgi funkcji eksponencjalnej w wyrażeniu (6.12) i przy założeniu prawdziwości hipotezy zerowej jest równa sumie kwadratów zmiennych losowych normalnych

$$\sqrt{\frac{n_l n_t}{n_l + n_t}} \frac{(\hat{x}_{kj}^{(l)} - \hat{x}_{kj}^{(t)})}{\sigma_{kj}}, \quad (6.14)$$

o zerowych wartościach oczekiwanych i jednostkowych wariancjach. Dlatego w przypadku prawdziwości hipotezy zerowej suma (6.13) będzie miała rozkład χ^2 o m stopniach swobody (Zołotariew, 1986).

Jeśli teraz zastąpimy w wyrażeniu (6.13) wielkość $D(\widetilde{X}_{ltj})$ jej estymatorem (6.10) a wielkość $\hat{x}_{ltj}^2$ przez sumę (6.9), to po uwzględnieniu równości (6.8) i wyrażenia (6.14) otrzymamy kryterium weryfikacji hipotezy o jednorodności grup, scharakteryzowanych za pomocą zbioru m cech. Po wykonaniu niezbędnych przekształceń kryterium to przyjmie postać:

$$\min_{l,t} U(K_l,K_t)=\min\frac{n_l+n_t-1}{n_l n_t(n_l+n_t)}\sum_{j=1}^{m}\frac{\left(n_t\sum_{k\in K_l}x_{kj}-n_l\sum_{k\in K_t}x_{kj}\right)^2}{\sum_{k\in K_l\cup K_t}x_{kj}^2-\frac{1}{n_l+n_t}\left(\sum_{k\in K_l\cup K_t}x_{kj}\right)^2}. \quad (5.15)$$

Jeśli $\min_{l,t} U(K_l,K_t)$ jest niewiększe od dopuszczalnej wartości statystyki $\chi^2_{\alpha,m}$, odpowiadającej z góry przyjętemu poziomowi istotności α oraz m stopniom swobody, to hipotezę o jednorodności uważa się za niesprzeczną z danymi empirycznymi i akceptuje się. Jeśli jednak $\min_{l,t} U(K_l,K_t)$ jest większe od $\chi^2_{\alpha,m}$, to hipotezę o jednorodności trzeba odrzucić.

Aby skorzystać z tej metody do zweryfikowania jednorodności zespołu złożonego z q odrębnych zbiorów, trzeba wykonać następujące czynności:

Krok 1

Na podstawie danych empirycznych trzeba utworzyć q macierzy danych postaci (4.35), przy czym macierz l $(l=1,2,...,m)$ l-tej grupie, jednorodnej w sensie jakiejś z góry ustalonej cechy. Liczba obserwacji w grupie l powinna spełniać warunek $n_l \geq 2$.

Krok 2

Dla każdej pary grup danych obliczamy wartość kryterium (5.15). Wyniki obliczeń przedstawiamy w postaci macierzy (z uwagi na symetryczność wypisujemy tylko elementy stojące nad główną przekątną).

$$\begin{pmatrix}
0 & U(K_1,K_2) & U(K_1,K_3) & \dots & U(K_1,K_t) & \dots & U(K_1,K_q)\\
 & 0 & U(K_2,K_3) & \dots & U(K_2,K_t) & \dots & U(K_2,K_q)\\
 & & 0 & \ddots & \vdots & & \vdots\\
 & & & 0 & U(K_l,K_t) & \dots & U(K_l,K_q)\\
 & & & & 0 & \ddots & \vdots\\
 & & & & & 0 & U(K_{q-1},K_q)\\
 & & & & & & 0
\end{pmatrix} \quad (5.16)$$

Krok 3

Spośród liczb wypełniających tę macierz wybieramy najmniejszą i porównujemy z dopuszczalną wartością statystyki $\chi^2_{\alpha,m}$. Jeżeli $\min_{l,t} U(K_l, K_t) > \chi^2_{\alpha,m}$, to wszystkie grupy uważa się za istotnie różniące się między sobą i procedura zostaje przerwana. Jeśli jednak $\min_{l,t} U(K_l, K_t) \le \chi^2_{\alpha,m}$, to parę (K_l, K_t), dla której kryterium (5.15) osiągnęło minimum łączy się w jedną grupę i całą procedurę weryfikacyjną powtarza się.

Przykład

Załóżmy, że obserwacja ocen kompetencji zbiorowości 58 nauczycieli pewnej szkoły wyższej z punktu widzenia pięciu cech (ocena merytoryczna - x_1, przygotowanie pedagogiczne - x_2, umiejętności dydaktyczne - x_3, umiejętności wychowawcze x_4, opinia środowiska studenckiego - x_5) pokazała, że procentowy udział każdej z nich w ogólnej ocenie nauczycieli był różny. Załóżmy, że skonstruowano ocenę całościową (np. suma ocen) i kierując się nią podzielono rozpatrywaną zbiorowość nauczycieli na osiem grup jednorodnych w sensie tej oceny. Następnie z każdej grupy wybrano losowo co najmniej dwóch nauczycieli i każdemu z nich przyporządkowano odpowiadający mu wektor ocen $(x_1, x_2, x_3, x_4, x_5)$. Otrzymano w ten sposób osiem hipotetycznie jednorodnych grup nauczycieli (danych), liczących, odpowiednio, 4, 5, 8, 9, 6, 11, 9 i 6 osób. Następnie, w odniesieniu do każdego z nauczycieli, obliczono procentowy udział każdej z rozpatrywanych pięciu ocen w ocenie całościowej. Wyniki obliczeń zostały przedstawione w postaci odpowiedniej macierzy danych. Dla zilustrowania struktury tej macierzy część danych jest przedstawiona w tablicy 6.1.

Tablica 6.1

Procentowy udział poszczególnych cech w całościowej ocenie nauczycieli (dla grup 1, 2 i 8)

Grupa	Kolejny numer nauczyciela	x_1	x_2	x_3	x_4	x_5
1	1	15.1	3.9	36.4	41.6	7.6
	2	15.6	3.4	33.0	42.3	6.7
	3	14.9	3.1	34.8	40.0	7.2
	4	18.3	3.7	28.7	43.4	7.9
2	5	10.2	3.1	32.8	45.9	7.9
	6	10.3	3.5	23.7	55.6	6.9
	7	14.1	2.9	26.4	48.4	8.2
	8	9.9	3.0	39.2	40.2	7.7

	9	19.1	3.3	29.8	36.7	11.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	52	12.8	2.9	29.8	46.4	8.1
	53	13.9	3.3	31.4	42.1	9.3
	54	11.9	1.8	28.9	49.5	7.8
	55	13.0	3.3	30.1	45.1	8.5
	56	17.2	3.5	28.6	39.0	11.7
	57	14.9	2.8	30.3	42.9	9.1
	58	14.9	3.4	31.1	40.8	9.8

Z porównania procentowych udziałów poszczególnych ocen w poszczególnych grupach wynika, iż można przyjąć, że grupa 1 jest w pewnym sensie wzorcowa. Przyjęcie tego postulatu rodzi jednak szereg pytań. W szczególności chcemy wiedzieć, czy wśród pozostałych grup istnieją grupy podobne do grupy 1, czy za pomocą metod statystycznych można podzielić rozpatrywaną zbiorowość nauczycieli na grupy o ustabilizowanej zmienności udziałów poszczególnych cech itp. Aby na te pytania odpowiedzieć trzeba przeprowadzić statystyczną analizę macierzy danych, posługując się kryterium (6.15). Procedura obliczeniowa ma następującą strukturę:

Krok 1

Obliczone wartości kryterium jednorodności $U(K_l, K_t)$ dla danych z tablicy 6.1 są przedstawione w tablicy 6.2. Najmniejszą wartością kryterium jest $U(K_4, K_5) = 1.36$. Odczytana z tablic statystycznych wartość $\chi^2_{0.05;5} = 11.07$ (poziom istotności $\alpha = 0.05$, liczba stopni swobody $m = 5$). Ponieważ $U(K_4, K_5) < \chi^2_{0.05;5}$, więc grupy 4 i 5 uznajemy za jednorodne. Łączymy je więc w jedną grupę i powtarzamy procedurę porównując otrzymaną grupę z sześcioma pozostałymi.

Tablica 6.2

Wartości statystyki $U(K_l, K_t)$ po pierwszym kroku

Grupa	2	3	4	5	6	7	8
1	8.91	8.54	6.99	8.78	10.23	14.61	4.23
2		5.01	5.13	3.02	9.87	12.21	6.47
3			3.47	2.01	10.45	9.56	2.98
4				1.36	7.99	9.98	3.14
5					5.68	10.11	3.57
6						8.92	2.79
7							9.12

Krok 2

Macierz nowych wartości kryterium $U(K_l, K_t)$ jest przedstawiona w tablicy 6.3. Najmniejsza wartość kryterium w grupie otrzymanej z połączenia grup 4 i 5 wynosi 2.98. Jest to wartość przeszło dwa razy większa od $U(K_4, K_5) = 1.36$. Połączenie grup 4 i 5 spowodowało wzrost wartości wariancji, a to z kolei spowodowało wzrost wartości kryterium jednorodności. Analiza pozostałych wartości kryterium pokazuje, że najmniejszą jego wartością jest $U(K_3, K_8) = 2.98$. Ponieważ wartość ta jest mniejsza od $\chi^2_{0.05;5} = 11.07$, więc grupy 3 i 8 można uznać za jednorodne i połączyć je ze sobą. Zauważmy, że większość wyników podanych w tablicy 6.2 nie zmieniła się. Nowe obliczenia trzeba było wykonać tylko dla sześciu par (ostatnia kolumna tablicy 6.3).

Tablica 6.3

Wartości statystyki $U(K_l, K_t)$ po drugim kroku

Grupa	2	3	6	7	8	$4 \cup 5$
1	8.91	8.54	10.23	14.61	4.23	8.99
2		5.01	9.87	12.21	6.47	8.04
3			10.45	9.56	2.98	3.01
6				8.92	3.02	9.73
7					9.12	10.02
8						4.98

Kroki 3, 4 i 5

Postępowanie opisane w Kroku 2 było kontynuowane w Krokach 3, 4 i5. Wyniki obliczeń są przedstawione w tablicach 6.4, 6.5 i 6.6.

Tablica 6.4

Wartości statystyki $U(K_l, K_t)$ po trzecim kroku

Grupa	2	6	7	$4 \cup 5$	$3 \cup 8$
1	8.91	10.23	14.61	8.99	3.54
2		9.87	12.21	8.04	5.07
6			8.92	9.73	8.01
7				10.02	9.99
$4 \cup 5$					0.08

Tablica 6.5

Wartości statystyki $U(K_l,K_t)$ po czwartym kroku

Grupa	2	6	7	$3 \cup 4 \cup 5 \cup 8$
1	8.91	10.23	14.61	67.23
2		9.87	12.21	[redacted]
6			8.92	12.27
7				11.43

Tablica 6.6

Wartości statystyki $U(K_l,K_t)$ po piątym kroku

Grupa	6	7	$2 \cup 3 \cup 4 \cup 5 \cup 8$
1	10.23	14.61	17.23
6		12.21	18.51
7			28.17

W tablicy 6.6 jest przedstawiony ostateczny wynik grupowania rozpatrywanej zbiorowości nauczycieli z punktu widzenia rozpatrywanych cech. Zamiast wyjściowych ośmiu grup otrzymano cztery grupy jednorodne w sensie statystycznym.

Krok 6

Wykorzystując dane przedstawione w tablicach 6.1 i 6.6 obliczono średnie wartości procentowego udziału każdej z rozpatrywanych cech w poszczególnych grupach. Wartości te są podane w tablicy 6.7.

Tablica 6.7

Średnie wartości procentowego udziału poszczególnych cech w poszczególnych grupach jednorodnych

Grupa	Liczba nauczycieli w grupie	x_1	x_2	x_3	x_4	x_5
1	4	28.1	36.2	23.6	5.8	6.3
6	11	26.4	34.6	26.4	8.9	4.5
7	9	27.4	37.3	22.7	6.7	5.9
$2 \cup 3 \cup 4 \cup 5 \cup 8$	34	26.9	36.7	21.1	9.1	6.2

Z tablicy 6.7 widać, że we wszystkich otrzymanych jednorodnych grupach nauczycieli największy udział w ocenie ma przygotowanie pedagogiczne (x_2). Drugim z kolei czynnikiem jest przygotowanie merytoryczne, a trzecim – umiejętności dydaktyczne. Następnym z kolei czynnikiem są umiejętności wychowawcze (x_4), a ostatnim – opinia środowiska studenckiego (x_5). Zwróćmy uwagę, że procentowy udział tych dwóch ostatnich czynników w każdej grupie jest dość istotny i wynosi średnio około 13%.

Otrzymane wyniki zwracają uwagę na to, jak duży wpływ na jakość procesu edukacyjnego w szkole wyższej mają przygotowanie pedagogiczne oraz umiejętności dydaktyczne nauczycieli. Potwierdza to znaną w środowiskach akademickich prawdę, że nie każdy naukowiec może być nauczycielem akademickim. Powinni nimi być tylko ci, którzy posiadają umiejętność przekazania wiedzy, umiejętność pracy z grupą i zmobilizowania studentów do aktywnego uczestnictwa w zajęciach podstawowych i dodatkowych (koła zainteresowań, uczestnictwo w ambitnych projektach itp.), oraz samokształcenia. Trzeba też zwrócić uwagę na to, że jakość zawodową takich nauczycieli można jeszcze bardziej podnieść przez odpowiednie przygotowanie pedagogiczne.

7. Metoda kojarzenia grup studenckich z nauczycielami

Przedstawione w poprzednich rozdziałach metody pozwalały podzielić zbiorowości studentów na grupy statystycznie jednorodne z punktu widzenia zespoły cech. Załóżmy, że wyodrębniono s grup studenckich oraz n grup nauczycielskich. Załóżmy też, że zarówno pierwsze, jak i drugie z nich są uszeregowane w kolejności od najlepszej do najsłabszej w sensie posiadanych kompetencji i umiejętności. Powstaje konieczność przyporządkowania każdej grupie studenckiej takiego nauczyciela, by łączny efekt procesu edukacyjnego był w danych warunkach jak najlepszy. Zadanie to na pierwszy rzut oka wydaje się banalne. Wiadomo przecież, że w przeszłości szkoły wyższe nieźle rozwiązywały to zadanie kierując się przede wszystkim zdrowym rozsądkiem, utrwaloną tradycją i doświadczeniem. W ostatnich dekadach konieczne stało się jednak wspomaganie tego procesu skomputeryzowanymi modelami matematycznymi, zwłaszcza w obszarze planowania i harmonogramowania rozkładów zajęć oraz zajętości pomieszczeń. Okazuje się, że to nie wystarcza. Coraz mocniej podkreśla się, że jakość edukacji w szkole wyższej zależy w znacznie większej mierze niż dotychczas sądzono od rozpoznania i racjonalnego wykorzystania umiejętności oraz kompetencji nauczycieli akademickich w procesie edukacyjnym (Hanoushek 1973; Perkins 1974; Connell 1980; McDonald 1980; Ravitch 1981; Soar, Medley i Coker 1983; Stark 1984; Kauchak, Peterson i Driscoll 1985; Tang 1988; Fullan 1992; Tate 1993; Elmore 1995; Hanoushek 1995; Oelkers 1996; Stoll i Fink 1996; Reid 1999; McAlpine i Weston 2000; Dorier 2001; Girod 2002; ICME 2004). Ponieważ umiejętności i kompetencje studentów i nauczycieli zmieniają się w czasie i muszą być oceniane w sposób wielowymiarowy, zadanie racjonalizacji przyporządkowywania nauczycieli grupom zajęciowym jest wielowymiarowym dynamicznym zadaniem decyzyjnym.

Istotnym znamieniem tego zadania jest niepełność i niepewność informacji charakteryzującej studentów i nauczycieli oraz nieokreśloność przyszłości, tym większa, im dłuższe są rozpatrywane horyzonty decyzyjne.

Zadanie racjonalnego przyporządkowania nauczycieli grupom studenckim może być sformułowane w różny sposób.

Niech I będzie liczbą grup studenckich, J liczbą nauczycieli. Załóżmy, że są dane atrybuty określające nauczyciela jako prowadzącego zajęcia z danego przedmiotu z określoną, odpowiednio wybraną, grupą studentów. Przyjmijmy, że studenci danego semestru zostali podzieleni na rozłączne jednorodne grupy według z góry ustalonych kryteriów. Kryteriami tymi mogą być, na przykład, wyniki uzyskane w poprzednim semestrze z wybranych przedmiotów kierunkowych. Oznaczmy przez a_s $(s = 1,2,...,S)$ atrybuty, według których podzielono studentów, a przez a_n $(n = 1,2,...,N)$ atrybuty, według których podzielono nauczycieli. Załóżmy, że wyróżniono l_s jednorodnych grup studenckich (oznaczmy je symbolami $G_1, G_2,..., G_{l_s}$) oraz l_n jednorodnych grup nauczycieli (oznaczmy je symbolami $F_1, F_2,..., F_{l_n}$).

Załóżmy, że istnieje funkcja V, której argumentami są atrybuty a_s i a_n, określająca użyteczność dla uczelni, wynikającą z przyporządkowania danego nauczyciela określonej grupie studenckiej. Można przyjąć, że przy właściwym przyporządkowaniu funkcja ta osiąga wartość maksymalną. Im gorsze jest to przyporządkowanie, tym mniejsza jest wartość funkcji V. W pierwszym przybliżeniu można przyjąć, że zależność ta ma charakter liniowy.

Oczywiście, najlepsze rozwiązanie uzyskujemy wtedy, gdy każdej grupie studenckiej $t = 1,2,...,l_s$ jest przyporządkowany taki nauczyciel, że funkcja V_v^t osiąga wartość maksymalną $V_{\max}^t$, gdzie $V_v^t(a_s, a_n)$ oznacza użyteczność uzyskaną w przypadku przyporządkowania nauczyciela z grupy F_v $(v = 1,2,...,l_n)$ do grupy studenckiej G_t $(t = 1,2,...,l_s)$. Można przyjąć, że wartości $V_{\max}^t$ są różne dla różnych grup G_t. Oczywiście w sytuacjach rzeczywistych takie przyporządkowanie nie jest możliwe. Można więc zaproponować, że przyporządkowanie nauczycieli do poszczególnych grup uzyskujemy w wyniku rozwiązania zadania optymalizacyjnego

$$\min \sum_{t=1}^{l_s} \sum_{v=1}^{l_n} (V_v^t - V_{\max}^t), \qquad (7.1)$$

przy ograniczeniu

$$V_v^t \geq V_{dop}^t, \tag{7.2}$$

gdzie V_{dop}^t określa próg, poniżej którego nie może spaść wartość funkcji użyteczności.

Podstawowym problemem przy takim podejściu pozostaje określenie postaci funkcji $V_v^t(a_s, a_n)$.

Jeżeli znane jest przyporządkowanie nauczycieli do grup studenckich, to dokonane przez osobę uznaną za eksperta, to stosując metody uczenia maszynowego można wyznaczyć reguły decyzyjne przyporządkowujące nauczycieli do grup studenckich. Reguły te można wykorzystać w kolejnych latach. W miarę zwiększania się danych dotyczących przyjętych uporządkowań reguły te mogą być uszczegółowiane.

8. Nierównowaga procesu edukacyjnego

8.1. Przyczyny i skutki braku równowagi

Proces edukacyjny polega na przeprowadzenia studenta (ucznia) ze stanu początkowego określonego jego zasobem wiedzy, przygotowaniem i motywacją do przyswajania nowej wiedzy do stanu końcowego wyznaczonego przez zbiór parametrów standardu danej dyscypliny i indywidualne aspiracje studenta. Plan studiów wyznacza tempo w jakim realizowane jest dochodzenie do stanu końcowego w przestrzeni wszystkich parametrów tego stanu.

Proces edukacyjny nazwiemy niezrównoważonym, jeśli tempo realizacji planu dla co najmniej dla jednego z parametrów jest nierównomierne w czasie. Trzema podstawowymi przyczynami niezrównoważenia są:

- niezrównoważony plan,
- trudności studentów (złe przygotowanie i trudności materialne),
- błędy nauczycieli i odstępstwa od planu.

Parametry standardów warunkujące plany studiów:

a) Punktowa miara nakładu pracy studenta – punkty ECTS
Ogólna liczba punktów ECTS,
Zakres tolerancji semestralnej liczby punktów ECTS,
Liczba punktów ECTS w grupach przedmiotów przewidzianych w standardzie:
przedmioty ogólne,
przedmioty matematyczne,
przedmioty specjalistyczne,
przedmioty inżynierskie,

lektoraty,
praca dyplomowa, itd.

	07/08 l	07/08 z	06/07 l	06/07 z	05/06 l	05/06 z	04/05 l	04/05 z	03/04 l	03/04 z	suma ECTS	rocznik
				28	30	32	30	30	31	29	180	2003
		30	28	30	30	32	31	29			180	2004
...	30	27	33	31	30	29					180	2005
...	30	30	30	30							180	2006
...	29	30									180	2007
...												

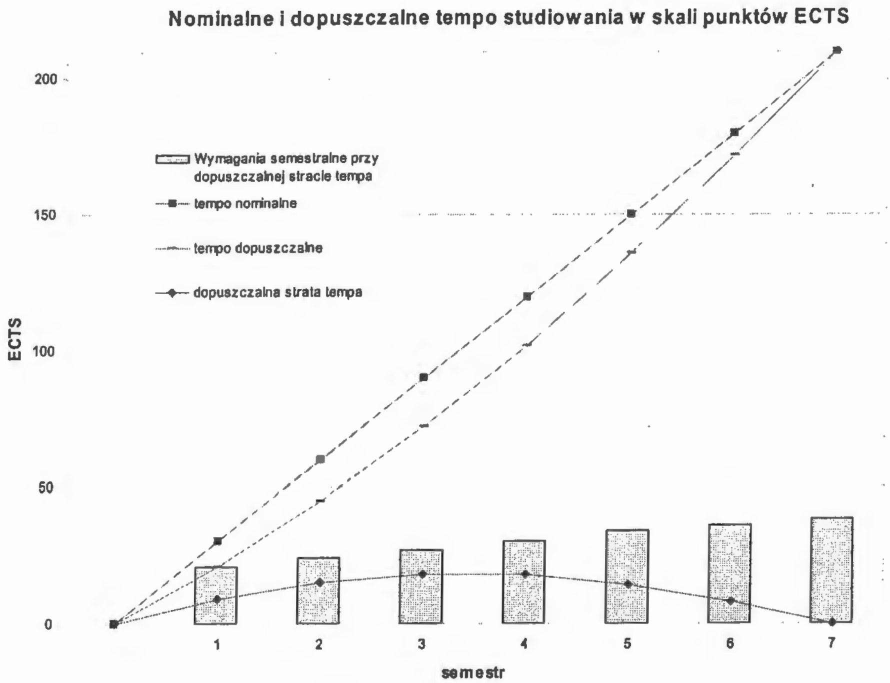
Rys. 8.1. Plany studiów i plany semestralne jako obciążenie wyrażone punktami ECTS

W wierszach tablicy przedstawionej na rys. 8.1. zapisane są plany studiów dla poszczególnych roczników studentów, w kolumnach – plany zajęć w kolejnych semestrach akademickich.

W praktyce zarówno planowe liczby punktów ETCS na poszczególnych semestrach jak ich akceptowalne realizacje w wynikach studentów zakładają określony dopuszczalny margines odstępstwa od założone tempa wzrostu. Należy jednak zadać sobie sprawę, że przy zadanych czasie procesu edukacyjnego (zadanej liczbie etapów) każde zwolnienie tempa w jednym etapie musi pociągać za sobą odpowiedni wzrost tempa w innych etapach. Może się więc zdażyć, że korzystanie z dopuszczalnego spadku tempa studiowania prowadzi do nierealizowalnego obciążenia w przyszłości. Na rys. 8.2 przedstawiono typową krzywą tolerancji stonowaną na wydziale przy podejmowaniu decyzji zaliczania studentom kolejnych etapów (semestrów). Student, który skorzysta z takiej tolerancji w pierwszej połowie studiów, z dużym prawdopodobieństwem nie ukończy studiów w planowanym terminie, gdyż jego obciążenie w ostatnich semestrach wzrasta o ponad 60% względem nominalnego.

b) Standardy godzinowe (wartości minimalne)
- ogólna liczba godzin dydaktycznych
- liczba godzin dydaktycznych w grupach przedmiotów,
- liczba godzin konkretnych przedmiotów (np. język obcy),
- liczba godzin zajęć w tygodniu (wartości maksymalne),
- dopuszczalne zmniejszenie liczby godzin dla studiów niestacjonarnych,
- dopuszczalna liczba godzin realizowanych w systemie zdalnym (distance learning).

c) Grupy następstwa przedmiotów w programie.

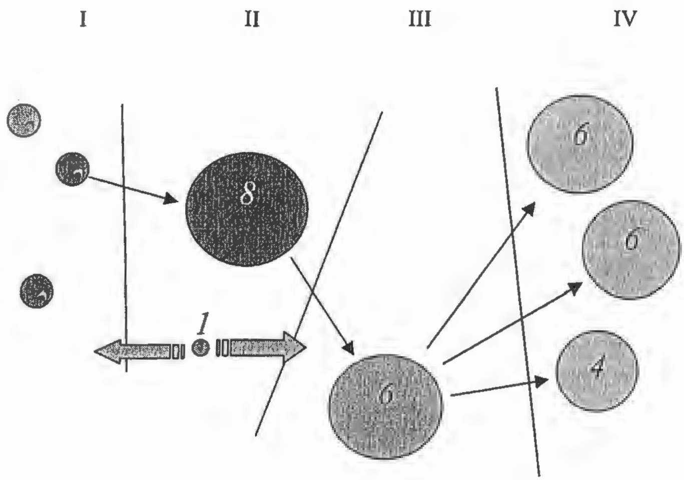


Rys. 8.2. Krzywa tolerancji

W wielu przypadkach przedmioty składające się na program studiów nie są jednostkami niezależnymi. Aby skutecznie podjąć studiowanie określonego materiału należy przed tym opanować wiedzę zawartą w innych przedmiotach – poprzednikach. Problem ten stanowi poważne wyzwanie przy projektowaniu planu studiów. Po pierwsze należy prawidłowo zdiagnozować merytoryczną potrzebę ustalenia relacji następstwa. Niewykrycie takiej relacji w treści programowej przedmiotów prowadzi do poważnego zakłócenia procesu edukacyjnego: studenci są źle przygotowani a nauczyciel musi poświęcić zbyt wiele czasu na uzupełnienie luk w ich przygotowaniu lub studenci nie są w stanie zaliczyć przedmiotu. Właściwie wykryte relacje następstwa ze swej strony stanowią istotne ograniczenie w konstruowaniu planu.

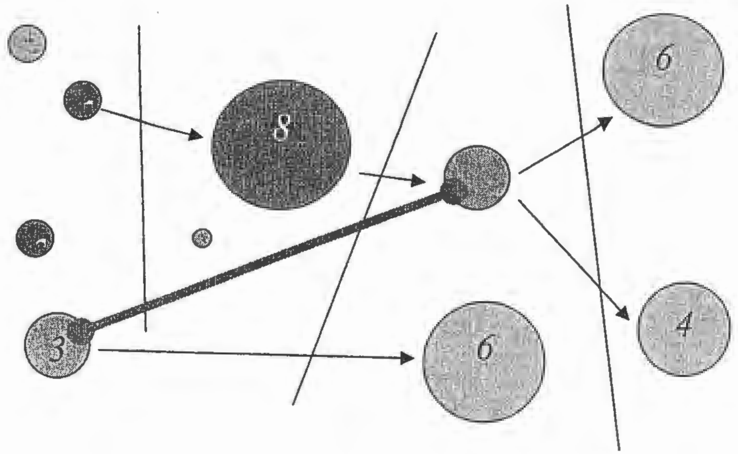
Formalnie dla zbioru przedmiotów relację następstwa można zapisać następująco:

- K przedmiotów programu dzielimy na $N \leq K$ grup tak, że do grupy i należy każdy przedmiot, który ma co najmniej jednego poprzednika obowiązkowego należącego do grupy i - 1.
- Przedmiot, który nie ma żadnego poprzednika obowiązkowego należy do grupy 1; musi istnieć co najmniej jeden taki przedmiot.
- W M – etapowym procesie edukacyjnym, w planie i – tego etapu mogą występować przedmioty z grup od 1 do i.
- Liczba etapów musi spełniać warunek $M \geq N$.



Rys. 8.3. Ilustracja konstrukcji planu studiów dla wielu różnych przedmiotów o niejednakowej pracochłonności wyrażone punktami ECTS.

W podanym przykładzie (rys. 8.3) pojawia się trudność związana z regułami następstw przedmiotów: nie ma możliwości utworzenia czteroetapowego procesu z zachowaniem zrównoważenia punktów ECTS w etapach; w czwartym etapie jest ich znacznie więcej niż w poprzednich – jedynym rozwiązaniem byłoby utworzenie piątego etapu.



Rys. 8.4. Konstrukcja planu po zmodyfikowaniu przedmiotów

Rys. 8.4 pokazuje jak, dekomponując jeden przedmiot programu, można zmodyfikować zadanie tak, aby czteroetapowy plan był zrównoważony. Korzystamy z tego, że część materiału pierwotnie ustalonego przedmiotu nie wymaga zaliczania poprzedników, a dodatkowo nowy przedmiot przesunięty do etapu I umożliwia zaliczenie jednego z przedmiotów poprzednio usytuowanych w etapie IV do etapu III.

8.2. Wskaźniki zrównoważenia realizacji planu (dla zbiorowości studentów)

Jak wspomniano na wstępie zrównoważenie realizacji planu zależy studentów i od nauczycieli akademickich. Aby właściwie diagnozować współdziałanie między nimi i wskazać nieprawidłowości stosuje się odpowiednie statystyki mające na celu wykrycie wspólnych cech całej populacji studentów, jak też testy pozwalające wyodrębnić grupy o jednorodnym zachowaniu w procesie edukacyjnym (patrz poprzednie rozdziały).

Poprawnie przeprowadzona analiza takich statystyk pozwala z jednej strony wskazać etapy i przedmioty, dla których istnieje zagrożenie dla realizacji planu z punktu widzenia całej populacji studentów, a drugiej wykryć niejednorodność populacji studentów z punktu widzenia realizacji planu w jego dynamicznym aspekcie.

W efekcie narzędzia opisane w poprzednich rozdziałach mogą być zastosowane także na wybranych etapach procesu edukacyjnego – tych, dla których powstaje zagrożenie zrównoważenia procesu. Może to być łatwiejsze i mniej kosztowne niż dobieranie grup studentów i wykładowców globalnie dla całego procesu.

Ważnymi zbiorami danych, charakteryzujących całą populację studentów, są:

- skumulowana średnia liczba punktów ECTS po kolejnych etapach,
- procent zaliczonych przedmiotów ogółem i w grupach przedmiotowych,
- możliwa do podjęcia lista przedmiotów ograniczona warunkami następstwa,
- średnia ocen w skali bezwzględnej (dla wszystkich przedmiotów i dla zaliczonych przedmiotów,
- pozycja rankingowa przedmiotów (według średniej ocen).

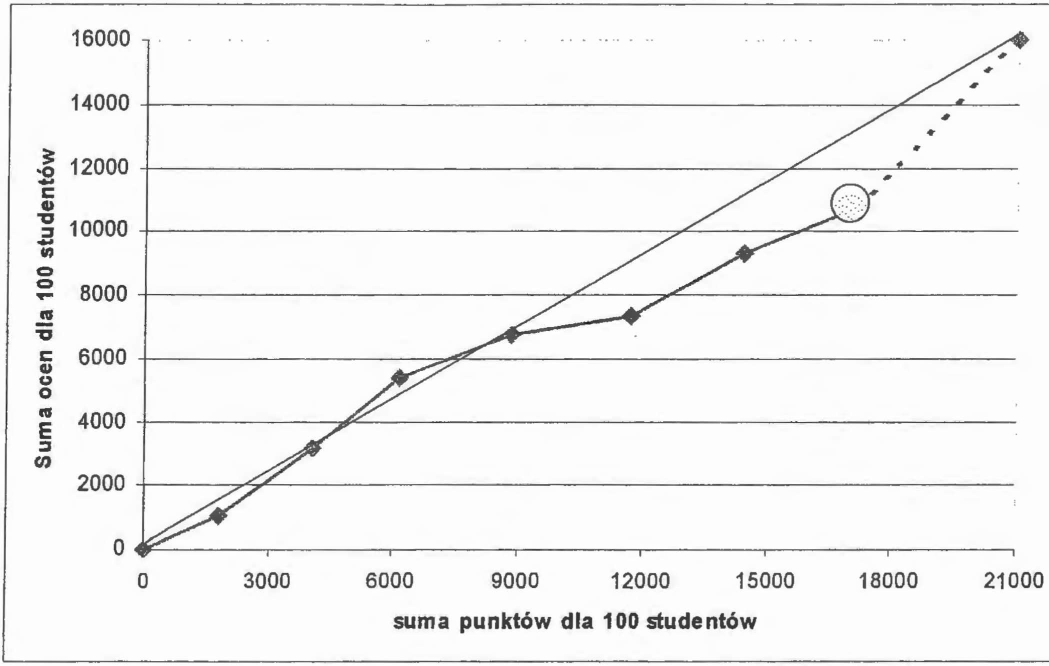
W celu wykrycia niejednorodności w populacji studentów tworzy się listy rankingowe średnich ocen studentów, średnich ocen z poszczególnych przedmiotów w kolejnych latach (kursów) oraz analizuje zmienność tych wielkości z etapu na etap.

Podstawowymi oznakami niezrównoważenia realizacji procesu dydaktycznego są:

- odbiegający od liniowego przebieg średniej uzyskanej skumulowanej liczby punktów ECTS w populacji studentów w kolejnych etapach (duży niedobór punktów lub duży procent nadwyżek),
- wysoki odsetek studentów, których niska skumulowana liczba punktów uniemożliwia awans na wyższy semestr,
- średnia ocen z przedmiotów zaliczonych skorelowana negatywnie z uzyskana liczba punktów ECTS (zaliczona mała liczba przedmiotów z wysoką średnią i nadwyżki punktów przy niskiej średniej.

Na rys. 8.5 przedstawiono trajektorię procesu edukacyjnego populacji 100 studentów w rzucie na płaszczyznę skumulowanych punktów ECTS i sumy uzyskanych ocen. Prawy górny narożnik wykresu reprezentuje idealny stan końcowy procesu: wszyscy studenci osiągnęli pełną liczbę punktów oraz wszyscy mieli wyłącznie oceny bardzo dobre (5).

Koło wskazuje punkt końcowy procesu w realnych warunkach: zarówno oceny są niższe jak i liczba uzyskanych punktów wskazuje, że nie wszyscy studenci uzyskali liczbę punktów umożliwiającą ukończenie studiów w ostatnim (siódmym) etapie. Odległość tego punktu od stanu idealnego może służyć jako miara jakości procesu.



Rys. 5 Przykład dwuwymiarowej analizy procesu edukacyjnego w przestrzeni: suma punktów ECTS – suma ocen

Równocześnie porównując kształt trajektorii z idealnym jej przebiegiem (odcinkiem prostej) można zauważyć, że w poszczególnych etapach niekorzystne zjawiska takie jak zahamowanie ogólnego przyrostu punktów ECTS lub niskie oceny (zwolniony przyrost ich sumy), zawsze negatywnie wpływające na położenie punktu reprezentującego stan końcowy, mają różne nasilenie w poszczególnych etapach. W przykładzie na rys. 8.5 słaby wynik końcowy procesu jest spowodowany w pierwszych trzech etapach małą liczbą uzyskanych punktów (wykorzystanie tolerancji przedstawionej na rys. 8.1, niezrównoważony plan lub nieprawidłowy wybór przedmiotów przez studentów) i w następnych etapach gorszymi ocenami, być może spowodowanymi przeciążeniem nadmiarem zajęć (nadrabianie opóźnień z pierwszych semestrów.

Zaproponowana powyżej metoda analizy zrównoważenia (lub jego braku) procesu edukacyjnego pozwala wykryć jego niejednorodność na osi czasu i dokonać alokacji jednostek programu i nauczycieli akademickich w celu zmniejszenia odległości stanu końcowego procesu od punktu idealnego (rys. 8.5).

9. Zakończenie i wnioski

U podstaw myśli dydaktycznej zawsze leżało założenie, że w procesie edukacyjnym występują dwa nierozdzielnie ze sobą powiązane podprocesy: proces nauczania i proces uczenia się (zob., np.: Nawroczyński 1957; Sośnicki 1959; Okoń 1995; Półturzycki 2002). Niestety, w badaniach procesów edukacyjnych prowadzonych przez środowiska nie związane bezpośrednio z edukacją z reguły analizowano te procesy niezależnie jeden od drugiego. Na przestrzeni kilku ostatnich dekad powstało wiele modeli matematycznych wyróżniających się nieraz formalną elegancją, lecz całkowicie sprzecznych z logiką procesu edukacyjnego. Poprawny model tego procesu musi uwzględniać systemową jedność procesu uczenia się (uczniowie, studenci) i nauczania (nauczyciele).

Praca dotyczy zastosowań metod statystycznych do oceny jakości procesu edukacyjnego. Tradycyjne metody oceniania traktują rozdzielnie proces nauczania i uczenia się. W pracy zwrócono uwagę na niesystemowość tego podejścia i przedstawiono koncepcję oceniania zintegrowanego, w którym uczestniczą obie strony procesu edukacyjnego: studenci i nauczyciele. U podstaw koncepcji leży sformułowany przez autorów postulat wieloaspektowej systemowej jednorodności grup zajęciowych oraz racjonalności kojarzenia odpowiednich nauczycieli akademickich z odpowiednimi grupami zajęciowymi. Przyjęto, że wśród cech charakteryzujących zbiorowości studentów bądź nauczycieli istnieje cecha priorytetowa z punktu widzenia celu badania. Sformułowano kryterium jednorodności. Ma ono postać wielowymiarowej statystyki, której rozkład jest tym bliższy rozkładowi chi-kwadrat, im liczniejsza jest rozpatrywana zbiorowość. Przedstawiono i przedyskutowano metodę dekompozycji niejednorodnej zbiorowości studentów (nauczycieli) na grupy jednorodne. Przeanalizowano rzeczywiste uwarunkowania procesu oceniania studentów oraz nauczycieli akademickich. Zwrócono uwagę na brak równowagi tego procesu i przedstawiono propozycję metody identyfikacji przyczyn i skutków tego zjawiska. Wskazano na możliwość poprawy jakości procesu edukacyjnego przez racjonalizację przyporządkowywania nauczycieli grupom zajęciowym. Sformułowano wstępną koncepcję systemowego kryterium jakości funkcjonowania szkoły wyższej, opartą na wykorzystaniu odpowiednio sformułowanej funkcji użyteczności, jako miary strat i ryzyka wynikających z niewłaściwego zorganizowania procesu edukacyjnego.

Wyniki pracy mają istotne znaczenie dla racjonalizacji procesu edukacyjnego, nie tylko w szkole wyższej. Przede wszystkim zwrócono uwagę na to, że grupy zajęciowe nie powinny być tworzone w sposób czysto administracyjny. Tworzenie grup jest bowiem wielowymiarowym dyskretnym zadaniem decyzyjnym. Grupy powinny być jednorodne w sensie statystycznym, tzn. w ich skład powinni wchodzić studenci, których poziom merytoryczny jest wprawdzie zróżnicowany, ale mieści się w ramach wyznaczonych przez wartości statystyki, stanowiącej kryterium jednorodności.

Ważnym wnioskiem jest również stwierdzenie, że jakość procesu edukacyjnego można istotnie podnieść wprowadzając dynamiczną organizację grup zajęciowych i dynamicznie optymalizując sposób kojarzenia tych grup z nauczycielami.

Literatura

Abbe E. (1863). Über die Gesetzmäßigkeit in der Verteilung der Fehler bei Beobachtungsreihen. W: Gesammelte Abhandlungen. Fischer, Jena.

Alexandersson H. (1986). A homogenaty test applied to precipitation data. *Journal of Climatology*, **6**, 661-675.

Anderson T. (1958). An introduction to multivariate statistical analysis. John Wiley, New York.

Anderson T.W. (1962). On the distribution of the two-sample Cramer – von Mises criterion. *Annals of Mathematical Statistics*, **33**, 1148-1159.

Anderson T.W., Darling D.A. (1952). Asymptotic theory of certain „Goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, **23**, 193-212.

Barra J.R. (1982). Matematyczne podstawy statystyki. Państwowe Wydawnictwo Naukowe, Warszawa.

Bartoszyński R., Niewiadomska-Bugaj M. (1996). Probabilisty and statistical inference. Jon Wiley, New York.

Bereziński M.(2003). Markowski model procesu edukacyjnego na wyższej uczelni. *Raport Badawczy RB/73/2003*, Instytut Badań Systemowych PAN, Warszawa.

Bereziński M., Inkielman M., Wagner D. (2004). Markowskie modele procesów edukacyjnych. *Raport Badawczy RB/63/2004*, Instytut Badań Systemowych PAN, Warszawa.

Bereziński M., Inkielman M., Wagner D. (2005a). Wieloaspektowy statystyczny model sterowania procesem edukacyjnym. *Raport Badawczy RB/45/2005*, Instytut Badań Systemowych PAN, Warszawa.

Bereziński M., Inkielman M., Wagner D. (2005b). Sieciowy stochastyczny model procesu kształcenia w szkole wyższej. W: Z. Bubnicki, R. Kulikowski i J. Kacprzyk, red., XV Krajowa Konferencja Automatyki, T. III. Instytut Badań Systemowych PAN, Warszawa, 359-364.

Bereziński M., Inkielman M., Wagner D. (2006). Metody oceny i doboru kadry dydaktycznej. *Raport Badawczy RB/43/2006*, Instytut Badań Systemowych PAN, Warszawa.

Bingham C., Nelson L.S. (1981). An approximation for the distribution of the von Neumann ratio. *Technometrics*, **23**, 285-288.

Bocheński J.M. (1993). Co to jest autorytet? W: Logika i filozofia. Państwowe Wydawnictwo Naukowe, Warszawa, 187-223.

Borowkow A.A. (1962). K zadaczie o dwóch wyborkach. *Izwiestia AN SSSR, Seria Matematiczeskaja*, **26**, 605-624.

Connell W.F. (1980). A history of education in the twentieth century world. Curriculum Development Centre, Canberra.

Cramér H. (1945). Mathematical methods of statistics. Almqvist and Wiksells, Uppsala.

Day N.E. (1959). Estimating the components of a mixture of normal distributions. *Biometrika*, **56**, 321-367.

Dorier J.-L. (2001). Panorama de mathématiques dans l'education française de la maternelle à l'université. Commission Française pour l'Enseignement des Mathématiques (CFEM), IREM, Paris.

Ebel R.L. (1980).Evaluation of students: Implications for effective teaching. *Educational Evaluation and Policy Analysis*, **2**, 47-51.

Elmore R.F. (1995). Structural reform in educational practice. *Education Researcher*, **24**, 23-36.

Feller W.(1969). Are life scientists overawed by statistics? *Scientific Research*, **4**, 24-29.

Fisher W.D. (1958). On grouping for maximum homogenity. *Journal of the American Statistical Association*, **53**, 789-798.

Fisz M. (1967). Rachunek prawdopodobieństwa i statystyka matematyczna. Państwowe Wydawnictwo Naukowe, Warszawa.

Fraser D.A.S. (1956). Nonparametric methods in statistics. John Wiley, New York.

Fullan M. (1992). Successful school improvement. The implementation perspective and beyond. Open University Press, Buckingham.

Girod G.R., red. (2002). Connecting teaching and learning: A handbook for teacher educators on teacher work sample. AACTE Publications, Washington.

Glasman N.S., Paulin P.J. (1982). Possible determinants of teacher respectivity to evaluation. *The Journal of Educational Administration*, **20**, 149-171.

Gnedenko B.V. (1976). The theory of probability. Mir Publishers, Moscow.

Guilford J.P. (1964). Podstawowe metody statystyczne w psychologii i pedagogice. Państwowe Wydawnictwo Naukowe, Warszawa.

Hald A. (1952). Statistical theory with engineering applications. John Wiley, New York.

Hanoushek E.A. (1973). Teacher characteristics and gains in student achievement: estimation using micro data. *The American Economic Review*, **61**, 280-288.

Hanoushek E.A. (1996). A more complete picture of school resource policies. *Review of EducationalResearch*, **66**, 397-409.

Hart B.I. (1942). Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **13**, 207-214.

Hellmert F.R. (1876). Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen. *Zeitschrift für Mathematik und Physik*, **21**, 192-218.

Hellwig Z. (1980). Elementy rachunku prawdopodobieństwa i statystyki matematycznej. Państwowe Wydawnictwo Naukowe, Warszawa.

Hotelling H. (1951). A generalized T-test and measure of multivariable dispersion. W: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probalbility. University of California Press, Berkeley, 23-41.

ICME (2004). Proceedings of the 10th International Congress on Mathematical Education (ICME 10, Copenhagen 4-11.07.2006)). European Mathematical Society Publishing House (EMS), Zürich, 2004.

Ito K. (1959). Asymptotic formulae for the distribution of Hotelling's generalized T^2 statistics. *Annals of Mathematical Statistics*, **27**, 1091-1105.

Kacprzyński BV. (1974). Planowanie eksperymentów. Podstawy matematyczne. Wydawnictwa Komunikacji i Łączności, Warszawa.

Kauchak D., Peterson K., Driscoll A. (1985). An interview study of teacher's attitudes toward teacher evaluation practices. *Journal of Research and Development in Education*, **19**, 32-37.

Kubik L. (1998). Zastosowanie elementarnego rachunku prawdopodobieństwa do wnioskowania statystycznego. Wydawnictwo Naukowe PWN, Warszawa.

Kupisiewicz C. (1984). Podstawy dydaktyki ogólnej. Państwowe Wydawnictwo Naukowe, Warszawa.

Linnik J.W. (1952). Metoda najmniejszych kwadratów i teoria opracowywania informacji. Państwowe Wydawnictwo Naukowe, Warszawa.

Marshall A.W. (1958). The small sample distribution of $n\omega_{n2}^2$. *Annals of Mathematical Statistics*, **29**, 307-309.

McAlpine, Weston C. (2000). Reflections: issues related to improving professors' teaching and student's learning. *Instructional Science*, **28**, 363-385.

McDonald F. (1980). Assessing teacher effects on student learning. W: D. Peterson i A. Ward, red., Due process in teacher evaluation. University Press of America, Lanham, 71-87.

Miller D.W., Starr M.K. (1969). Praktyka i teoria decyzji. Państwowe Wydawnictwo Naukowe, Warszawa.

Nawroczyński B. (1957). Zasady nauczania. Ossolineum, Wrocław.

Neumann J., von (1941). Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **12**, 367-395.

Neumann J., von (1942). A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **13**, 86-88.

Neumann J., von, Kent R.H., Bellinson H.R., Hart B.I. (1941). The mean square successive difference. *Annals of Mathematical Statistics*, **12**, 153-162.

Niemierko B. (1975). Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe. Wydawnictwa Szkolne i Pedagogiczne, Warszawa.
Niemierko B. (1990). Pomiar sprawdzający w dydaktyce. Państwowe Wydawnictwo Naukowe, Warszawa.
Niemierko B. (1991). Między oceną szkolną a dydaktyką. Bliżej dydaktyki. Wydawnictwa Szkolne i Pedagogiczne, Warszawa.
Niemierko B. (1999). Pomiar wyników kształcenia. Wydawnictwa Szkolne i Pedagogiczne, Warszawa.
Neumann J., von (1941). Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **12**, 367-395.
Neumann J., von (1942). A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics*, **13**, 86-88.
Oelkers J. (1996). Reformpädagogik: Eine kritische Dogmengeschichte. Weinheim, Juventa.
Okoń W. (1995). Wprowadzenie do dydaktyki ogólnej. Wydawnictwo Żak, Warszawa.
Oktaba W. (1966). Elementy statystyki matematycznej i metodyka doświadczalnictwa. Państwowe Wydawnictwo Naukowe, Warszawa.
Perkins J.A. (1974). Pięć przyczyn kryzysu uniwersytetu. *Dialogue – USA*, **3**, 3-8.
Półturzycki J. (2002). Dydaktyka dla nauczycieli. Wydawnictwo Naukowe NOVUM, Płock.
Rao C.R. (1994). Statystyka i prawda. Wydawnictwo Naukowe PWN, Warszawa.
Rasiowa H. (1979). Wstęp do matematyki współczesnej. Państwowe Wydawnictwo Naukowe, Warszawa. Rosenblatt M. (1952). Limit theorems associated with variants of the von Mises statistics. *Annals of Mathematical Statistics*, **23**, 617-623.
Ravitch D. (1981). Some answers ... and new questions. *The American Scholar*, **50**, nr 3.
Řehak J., Řehakova B. (1973). Měřěni statisticke závilosti nominálnich znaku. *Sociologicky Časopis*, **4**, 404-417.
Reid J. (1999). Improving teaching in higher education: student and teacher perspectives. *Educational Science*, **24**, 269-282.
Sidorowicz A. (1996). Brońmy autorytetu nauczyciela. *Edukacja i Dialog*, nr 7 6-7.
Smirnow N.W. (1937). O raspriedielenii ω^2 - kriteria Misesa. *Matematiczeskij Sbornik*, **44**, 973-994.
Smirnow N.W. (1944). Prybliżenije zakonow raspriedielenija słuczajnych wieliczin po empiriczeskim dannym. *Uspiechi Matematiczeskich Nauk*, **10**, 179-206.
Soar R.S., Medley D.M., Coker H. (1983). Teacher evaluation: A critique of currently used methods. *Phi Delta Kappan*, **64**, 239-246.
Sośnicki K. (1958). Dydaktyka ogólna. Ossolineum, Wrocław.
Stark J.S., Lowther M.A. (1984). Predictors of teacher's preference concerning their evaluation. *Educational Administration Quarterly*, **20**, 76-106.

Stoll L., Fink D. (1996). Changing our schools. Open University Press, Buckingham.

Tang M.C. (1988). Cost analysis for educational policymaking: A review of cost studies in education in developing countries. *Review of Educational Research*, **57**, 181-230.

Tate P. (1993). The two worlds of teaching. *Journal of Education*, **175**, 15-30.

Tatubalin W.N. (1977). Granicy primienimosti. Znanije, Moskwa.

Willams J.D. (1941). Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe. *Annals of Mathematical Statistics*, **12**, 239-241.

Young L.C. (1941). On randomness indered sequences. *Annals of Mathematical Statistics*, **12**, 293-300.

Zołotariew W.M. (1986). Sowremiennaja teoria summirowanija słuczajnych wieliczin. Nauka, Moskwa.

Spis treści