

Boris Kožuh

STOSOWANIE I INTERPRETACJA TESTÓW HIPOTEZ STATYSTYCZNYCH

Wstęp

W ostatnich dziesięcioleciach można zauważyć wielki rozwój badań empirycznych, zarówno w kraju, jak i na świecie. Opiera się dziś na nich większość rozpraw doktorskich. Podobny trend można zauważyć również w pracach magisterskich. Dominują one na studiach humanistycznych, zwłaszcza wśród studentów kierunków pedagogicznych.

Opracowania statystyczne w tych badaniach przeważnie zawierają różne procedury weryfikacji hipotez statystycznych – zazwyczaj hipotez zerowych. Wśród nich najczęściej wykorzystuje się chi-kwadrat test, a nieco rzadziej znajdują zastosowanie analiza wariancji i kowariancji, z-test, t-test oraz F-test. Pozostałe testy stosuje się sporadycznie.

Raporty i interpretacja wymienionych testów mogą być odmienne w różnych sytuacjach badawczych. Regułę, która zakłada, że służą one jedynie do wnioskowania, należy w czasach współczesnych skorygować i nieco dopełnić. Od czasu, gdy tak silnie rozwinęły się analizy wtórne, a szczególnie metaanaliza, pojawia się potrzeba sporządzania bardziej obszernych i pełniejszych raportów z zastosowanych testów. Ich pełne wersje należy opracować przynajmniej dla najważniejszych pytań badawczych, nawet w wypadku, gdy hipotezy zerowe nie zostały odrzucone.

Zamiarem autora jest pokazanie niektórych praw dotyczących stosowania oraz interpretacji najczęściej stosowanych testów istotności. W analizie tej zostanie uwypuklona potrzeba dokonywania analiz wtórnych i badań metaanalitycznych.

Hipotezy badawcze

Prawie każde badanie, a zwłaszcza empiryczne, zawiera hipotezy. Stanowią one niezwyfikowane odpowiedzi na problem badawczy lub na szczegółowe pytanie badawcze – „zbudowanie hipotez jest zabiegiem polegającym na zbudowaniu domniemanej teorii dotyczącej natury zjawiska, powiązań między jego elementami, jego wielkością, proporcją itp.” [Pilch, Bauman, 2001, s. 47]. Hipotezy opierają się najczęściej na dotychczasowej teorii. W klasycznym modelu badawczym na początku definiuje się ogólne pytanie badawcze lub problem badawczy. Etap ten jest często najbardziej czasochłonny, bowiem wymaga długiej kwerendy literatury. Jednak – dobrze wykonany – oznacza wielki krok w przebiegu badania. [Pilch, Bauman, 2001, s. 190-191]. W momencie, gdy zostało zdefiniowane pytanie ogólne (problem badawczy), można z góry założyć odpowiedź na nie. Tę odpowiedź nazywamy hipotezą. W związku z tym, że najważniejszymi prawami pedagogiki są te, które dotyczą relacji i związków między zjawiskami edukacyjnymi, to najczęściej też zarówno pytania badawcze, jak i hipotezy badawcze, dotyczą tychże relacji i związków. Zdecydowanie rzadziej hipotezy mówią o stanie rzeczy w badanej populacji (np. ilu nauczycieli korzysta z komputera przy przygotowaniu się do lekcji, ilu uczniów posiada swoje biurko itd.).

Do zagadnienia formułowania hipotez trzeba jednak podejść z pewną ostrożnością, a nawet z wątpliwościami i pytaniem: jaki jest sens formułowania hipotez, gdy odpowiedź, którą proponuje hipoteza, jest przedmiotem poszukiwania (czyli celem badania)? Wszystkie działania w badaniu są ukierunkowane na znalezienie odpowiedzi na pytanie badawcze. Warto zatem zastanowić się, po co stawia się hipotezę badawczą na początku badania, jeżeli ma ono dopiero znaleźć odpowiedź?

Sens hipotez zawiera się w fakcie, że powinny one ukierunkować badanie ku istotnym elementom i aspektom badanego problemu. Brak hipotezy prowadziłyby w konsekwencji do sytuacji, w której w każdym badaniu należałoby startować od zera – tak, jakby badacz nie wiedział zupełnie nic o badanym problemie lub tak, jak gdyby nie istniały żadne bardziej ogólne teorie, dotyczące badanych problemów.

W bardziej ilościowych badaniach hipotezy powstają z reguły w sposób dedukcyjny, czyli na początku badania. Natomiast w badaniach bardziej jakościowych hipotezy częściej tworzy się w sposób indukcyjny – w trakcie samego badania [Pilch, Bauman, 2001, s. 278]. Słabym, a nawet złym rozwiązaniem okazuje się sytuacja, w której badacz nieelastycznie kieruje się jedynie hipotezami stawianymi za pomocą dedukcji. Równie niepoprawna pozostaje też orientacja badacza wyłącznie na hipotezy pochodzące z indukcji.

Poszukiwanie odpowiedzi na pytania badawcze oznacza w istocie weryfikację hipotez. Jednak najczęściej pojęcie „testowania hipotez” stosuje się w wypadku weryfikacji statystycznej (np. hipotez zerowych). Weryfikacja następuje z wykorzystaniem różnych metod (nie tylko przez testowanie hipotez statystycznych). Upraszczając ten wywód, można stwierdzić, że wszystkie działania w badaniu służą ustaleniu prawdziwości przyjętych hipotez.

Hipotezy statystyczne

Badania sondażowe w obszarze edukacji zazwyczaj prowadzi się na próbach. Oznacza to, że bada się w istocie całą populację, ale materiał empiryczny jest gromadzony tylko w odniesieniu do próby badawczej (np. poprzez obserwację, ankiety, różne testy, skale itd.). Na podstawie danych z próby wyciąga się wnioski, które odnoszą się do całej populacji. W takim modelu badań, oprócz hipotez badawczych, pojawiają się także hipotezy statystyczne. Powstają one na etapie statystycznego opracowywania zgromadzonych danych i służą jako podstawa oraz niezbędny warunek przeprowadzenia procedur statystycznych. Dzięki hipotezom statystycznym można dokonać weryfikacji hipotez badawczych.

W celu jak najbardziej rzetelnej i pewnej weryfikacji hipotez badawczych z zebranego materiału badawczego należy zastosować precyzyjne metody statystyczne. Badacze nie powinni kierować się swoim przeczuciem, oceniając subiektywnie, czy wyniki potwierdzają, czy też nie potwierdzają założonej hipotezy. Z tego powodu weryfikacja hipotez badawczych nie jest najczęściej możliwa bez weryfikacji hipotez statystycznych. Należy jednak wyraźnie podkreślić, iż ta reguła dotyczy tylko badań na próbach, czyli przypadków, gdy generalizuje się wyniki uzyskane z próby na populację generalną. Zarówno hipotezy badawcze, jak i hipotezy statystyczne, dotyczą zawsze populacji generalnych.

W badaniach często hipotezy statystyczne okazują się specyficzne, hipotezy badawcze natomiast brzmią bardziej ogólnie. Reguła ta dotyczy jednak raczej strony formalnej, zaś istota relacji pozostaje inna. W rzeczywistości – przynajmniej w rozważaniach teoretycznych – każdą bardziej ogólną hipotezę badawczą należy rozłożyć na hipotezy częściowe – bardziej specyficzne (podobnie jak wcześniej pytania badawcze). Dzięki temu hipotezy zawsze występują w parach: hipoteza badawcza i hipoteza statystyczna. Takich par w badaniu występuje na ogół więcej – kilkanaście, a nawet kilkadziesiąt.

Relacja pomiędzy hipotezami badawczymi i statystycznymi

Hipoteza statystyczna jest zawsze przeciwstawna hipotezie badawczej. Relacje pomiędzy nimi (zarówno w aspekcie „ogólnie – specyficznie”, jak i „twierdzenie – przeczenie”) zostaną zilustrowane przykładem z praktyki. W badaniu sukcesów szkolnych uczniów gimnazjum porównywano dzieci, które mieszkają w miejscowości, gdzie jest szkoła, z tymi, które do szkoły muszą dojeżdżać. Jedno z pytań badawczych brzmiało: czy sukcesy szkolne tych uczniów różnią się? Ogólna hipoteza brzmiała: tak, sukcesy szkolne różnią się, ponieważ większe sukcesy odnoszą uczniowie, którzy nie dojeżdżają do szkoły. Hipoteza ta opiera się na fakcie, że uczniowie dojeżdżający zużywają dużo czasu i swojej energii podczas codziennych podróży do i ze szkoły, mają mniej kontaktów z innymi uczniami z klasy, trudniej jest im uczestniczyć w zajęciach pozalekcyjnych itd. Przy empirycznej weryfikacji tej hipotezy zmienną określoną jako „sukces szkolny” zbadano za pomocą wszystkich ocen otrzymanych przez uczniów w roku szkolnym. W ten sposób ogólna hi-

hipoteza została podzielona na szereg hipotez specyficznych (np. istnieją różnice ocen z języka polskiego, z matematyki itd.). W badaniu za wskaźniki zmiennej „sukces szkolny” można uznać tylko niektóre oceny, a nawet tylko średnią ocen – założenia w każdym badaniu okazują się bowiem inne. Jedną z hipotez specyficznych brzmiała: średnia ocen uczniów, którzy dojeżdżają do szkoły, jest niższa niż średnia ocen uczniów, którzy nie muszą dojeżdżać. Hipoteza statystyczna brzmiała: średnia ocen obydwu grup nie różni się. Hipotezy statystyczne prawie zawsze występują w takiej postaci i zakładają, iż nie ma różnicy (nie ma też współzależności itd.). Są to hipotezy zerowe.

Weryfikacja hipotez statystycznych

Pierwszy krok na tym etapie stanowi zgromadzenie danych w próbie. W przypadku, gdy badacz interesuje, czy w próbie średnia ocen obydwu grup różni się, wystarczy wyliczyć średnie ocen i porównać je między sobą. Jeżeli występuje różnica, to jest to już odpowiedź na pytanie badawcze. Nie potrzeba żadnej hipotezy statystycznej, tak samo jak niepotrzebny okazuje się jakikolwiek dodatkowy test. Podobnie jest w sytuacji, gdy badacz dysponuje danymi dla całej populacji (populacji generalnej). Proste porównanie dwóch średnich ocen wystarcza wówczas do udzielenia odpowiedzi na pytanie badawcze. Różnica tych średnich sama w sobie jest dowodem. W tym konkretnym wypadku należy zwrócić uwagę na problem, który powstaje przy stosowaniu oraz interpretacji średniej arytmetycznej dla ocen. Oceny szkolne to zmienna porządkowa (pomimo, że wartości liczbowe stwarzają złudzenie, iż jest to zmienna przedziałowa). Oznacza to, że przedziały pomiędzy ocenami nie są jednakowe (np. różnica pomiędzy oceną niedostateczną i dopuszczającą nie jest taka sama, jak pomiędzy oceną dobrą i bardzo dobrą itd.). Stąd sumowanie ocen nie jest uzasadnione – czyli nie można wyliczyć średniej arytmetycznej. Natura tej zmiennej pozwala jednak na zastosowanie na przykład mediany, choć bardzo często w wypadku ocen szkolnych prowadzi ono donikąd, ponieważ mediany – nawet dla ewidentnie różniących się grup – rzadko bywają zróżnicowane. Dlatego też jedynym rozwiązaniem pozostaje często zastosowanie średniej arytmetycznej. W takich sytuacjach należy bardzo ostrożnie interpretować wyniki – trzeba dobrze przemyśleć, co średnie arytmetyczne pokazują w rzeczywistości [zob. Ferguson, Takane, 1997, s. 69-80].

Całkowicie odmienna sytuacja występuje, kiedy zarówno pytania badawcze, jak i hipotezy dotyczą całej populacji generalnej, a zgromadzone dane pochodzą tylko z próby. Są to jedyne sytuacje, w których potrzebne i sensowne jest zastosowanie statystycznych testów weryfikacji hipotez. Stawia się wówczas – oprócz hipotezy badawczej – również hipotezę statystyczną i stosuje się testy dla jej weryfikacji. W hipotezach badawczych najczęściej zdarza się, że wyraz „populacja generalna” wcale nie występuje, bo jest to oczywiste. Natomiast w przypadku hipotez statystycznych zawsze wyraźnie zaznacza się, że dotyczą one populacji generalnych. Dlaczego te hipotezy nigdy nie dotyczą prób? Dlatego, że o próbach wszystko już wiadomo, gdyż zebrano potrzebne dane na ich temat i nie ma potrzeby stawiania hipotez.

Sytuacja, w której średnie arytmetyczne prób różnią się od siebie, nie oznacza jeszcze i nie dowodzi, że średnie arytmetyczne populacji generalnych także się róż-

nią. Dotyczy to również wszystkich różnic pomiędzy próbami (a na ogół dotyczy to wszystkich wniosków na podstawie danych z prób). Dowodów o populacji generalnej może dostarczyć dopiero test weryfikacyjny.

Chi-kwadrat test

Test ten stosuje się w badaniach edukacji najczęściej. Jego popularność wynika stąd, że dane o zjawiskach edukacyjnych w większości należą do zmiennych opisowych (nominalnych i porządkowych). Test ten stosuje się głównie wtedy, gdy bada się współzależności pomiędzy zjawiskami. W takich przypadkach hipoteza badawcza zakłada zwykle, że zjawiska są powiązane, a hipoteza statystyczna natomiast, że zjawiska te w populacji generalnej są niezależne (hipoteza niezależności). Należy jednak podkreślić, że test chi-kwadrat nie służy tylko do weryfikowania hipotez niezależności – istnieje szereg innych sytuacji, w których można zastosować tę procedurę. W dalszej części tej rozprawy zostaną przedstawione dwa charakterystyczne przypadki stosowania oraz interpretacji testu chi-kwadrat.

Badanie nr 1

W tym badaniu ustalano współzależność pomiędzy wiekiem respondentów a ich motywacją do podjęcia studiów niestacjonarnych. Pytanie dotyczyło konkretnie tego, czy wiek ma wpływ na motywację do podjęcia nauki, a jeżeli ma wpływ, to jaki on jest. Pytanie to odnosiło się oczywiście do populacji generalnej. Hipoteza badawcza brzmiała: wiek ma wpływ na motywację do podjęcia studiów niestacjonarnych – u młodszych respondentów motywacja ta okazuje się silniejsza. Dane zostały zebrane w próbie. Postawiono hipotezę niezależności: motywacja do podjęcia nauki na studiach niestacjonarnych w populacji generalnej nie zależy od wieku.

Tabela 1. Odpowiedzi respondentów według wieku i motywacji do podjęcia nauki na studiach niestacjonarnych

wiek	motywacja			
	mała	średnia	duża	razem
do 25 lat	16 (25,4%)	20 (31,7%)	27 (42,9%)	63 (100,0%)
25-30 lat	12 (31,6%)	16 (42,1%)	10 (26,3%)	38 (100,0%)
ponad 35 lat	31 (58,5%)	14 (26,4%)	8 (15,1%)	53 (100,0%)
razem	59 (38,3%)	50 (32,5%)	45 (29,2%)	154 (100,0%)

$$\chi^2 = 18,26 > \chi^2 (\alpha=0,01; df=4) = 13,28$$

$$C = 0,33$$

Źródło: badania własne

Obliczona wartość chi-kwadrat przewyższa wartość wymaganą dla istotności na poziomie 0,01, dlatego też hipotezę niezależności odrzuca się z ryzykiem 1%. Jest to dowód, że w populacji generalnej wiek ma wpływ na motywację do podjęcia nauki na studiach niestacjonarnych. Należy jednak wyraźnie podkreślić, że dowód ten obciąża ryzyko wynoszące 1%.

W badaniu należy jeszcze wyjaśnić kierunek współzależności. Rozkład procentowy wskazuje, że respondenci w młodszym wieku wyrażają większe zainteresowanie podjęciem studiów. Analiza dowodzi także, iż motywacja od najmłodszej do najstarszej kategorii wiekowej systematycznie maleje. Po otrzymaniu wyników badacz powinien podjąć trud interpretacji i poszukiwania odpowiedzi na pytanie o przyczyny takiego stanu. Dodatkowo można obliczyć stopień współzależności za pomocą współczynników zbieżności. Z danych umieszczonych w tabeli obliczono współczynnik zbieżności Pearsona. Otrzymano wynik o wartości 0,33. Uwzględniając naturę zmiennych, można ten rezultat zinterpretować jako stosunkowo ciasną współzależność. Obliczanie współczynników jest potrzebne przede wszystkim wtedy, gdy porównuje się różne tabele lub własne wyniki z wynikami innych, wcześniejszych badań.

Podobnie, jak w wyżej opisany sposób postępuje się w sytuacji badawczej, gdy wartość chi-kwadrat przewyższa wartości krytyczne z tabeli na poziomie $P=0,05$ lub $P=0,001$. W obydwu wariantach hipoteza niezależności powinna zostać odrzucona, tylko na innym poziomie i z innym ryzykiem.

Badanie nr 2

W badaniu ustalano współzależność pomiędzy wykształceniem rodziców a ich współpracą z nauczycielami. Główne pytanie badawcze brzmiało: czy wykształcenie rodziców ma wpływ na ich współpracę z nauczycielami? Przyjęta hipoteza badawcza zakładała, że wykształcenie ma rzeczywiste znaczenie i bardziej wykształceni rodzice lepiej współpracują z nauczycielami. Poniższa tabela zawiera dane dotyczące wymienionych zmiennych pochodzących z próby. Przyjęto następującą hipotezę niezależności: współpraca z nauczycielami i wykształcenie są zmiennymi niezależnymi (oczywiście w populacji generalnej).

Tabela 2. Odpowiedzi respondentów według wykształcenia i częstotliwości kontaktów z wychowawcą klasy dziecka

	często	czasami	bardzo rzadko	razem
wykształcenie podstawowe	14 (24,6%)	19 (33,3%)	24 (42,1%)	57 (100,0%)
wykształcenie średnie	17 (43,6%)	10 (25,6%)	12 (30,8%)	39 (100,0%)
wykształcenie licencjackie	22 (45,8%)	18 (37,5%)	8 (16,7%)	48 (100,0%)
wykształcenie magisterskie	11 (45,8%)	6 (25,0%)	7 (29,2%)	24 (100,0%)
razem	64 (38,1%)	53 (31,5%)	51 (30,4%)	168 (100,0%)

$$\chi^2 = 11,10 < \chi^2 (\alpha=0,05; df=6) = 12,59$$

Źródło: badania własne

Obliczona wartość chi-kwadrat okazuje się mniejsza niż wartość wymagana dla istotności na poziomie 0,05, dlatego też nie należy odrzucać hipotezy niezależności. Taki wynik testu nie pozwala na żadne uogólnienie dla populacji generalnej. Oznacza to, iż nie istnieje dowód o zależności wymienionych zmiennych w populacji generalnej. Narzuca się następujący wniosek: jeżeli nie ma dowodu o zależności, oznacza to, że zmienne są niezależne. Wniosek ten jest jednak niepoprawny. Test chi-kwadrat nie może potwierdzać hipotezy niezależności. Dlatego też w takich sytuacjach wnioskuje się, że test niczego nie dowodzi i że o populacji generalnej nie można niczego stwierdzić na jego podstawie. To nieco niebezpieczny punkt całego testu, ponieważ w praktyce badawczej stosunkowo często pojawiają się niepoprawne interpretacje. Często błąd ten występuje w bardziej niewinnej formie, na przykład jeżeli hipoteza badawcza zakładała, że zmienne są zależne, to fakt, że hipoteza niezależności nie została odrzucona, nie pozwala na stwierdzenie niepoprawności hipotezy badawczej. W takiej sytuacji o populacji generalnej nie da się niczego powiedzieć. Procenty w tabeli, które przecież świadczą o jakichś trendach, interpretuje się tak samo, jak w wypadku pierwszego badania, lecz z wyraźnym zaznaczeniem, że cała interpretacja dotyczy tylko i wyłącznie próby. Naukowa wartość nowych ustaleń okazuje się wówczas zdecydowanie mniejsza, ale nie traci całkowicie znaczenia. Jeśli zaś wartość chi-kwadrat znajduje się na przykład nieco pod granicą istotności, stanowi to wskazówkę, w jakim kierunku warto pogłębić badanie, by ewentualnie później odnaleźć reguły dotyczące populacji generalnej.

Raport testu

Raport testu powinien dostarczyć czytelnikowi wszystkich informacji potrzebnych do poprawnego zrozumienia interpretacji. Pod wpływem częstego stosowania analiz wtórnych (szczególnie metaanalizy) w ostatnim dziesięcioleciu niektóre zasady tworzenia raportów uległy zmianie. Oznacza to, że raporty rozszerzyły się o te elementy, które są potrzebne do późniejszych metaanaliz. Przyjmuje się jako minimum, że takie zestawienie powinno zawierać następujące elementy: wartość chi-kwadrat, stopnie swobody oraz poziom istotności. Te elementy powinny znaleźć się w raporcie nawet wtedy, gdy hipoteza zerowa nie została odrzucona. Jednak – ponieważ obecnie już tylko incydentalnie opracowania statystyczne wykonuje się „ręcznie”, bez zastosowania komputera oraz programów statystycznych – pominę w tym artykule szczegółowe wyjaśnienia dotyczące starej formy raportu, ograniczając się jedynie do podstawowego porównania.

Stary sposób (za pomocą tabel wartości krytycznych):

$$\chi^2 = 11,10 < \chi^2 (\alpha = 0,05; df=6) = 12,59$$

$$\chi^2 = 18,26 > \chi^2 (\alpha = 0,01; df=4) = 13,28$$

W tej procedurze opierano się na trzech standardowych wartościach krytycznych: z ryzykiem 5%, z ryzykiem 1% i z ryzykiem 0,1% (i na odpowiednich poziomach istotności 0,05; 0,01 i 0,001).

Nowy sposób (zapis z komputera)

W opracowaniu komputerowym niepotrzebna okazuje się tabela wartości krytycznych, ponieważ program sam wylicza poziom istotności za pomocą równania krzywej chi-kwadrat. Oto kilka charakterystycznych przykładów z praktyki badawczej autora:

wartość χ^2	stopnie swobody (df)	poziom istotności (α)
32,65	20	0,0369
40,64	20	0,0041
12,02	15	0,6777
56,50	20	0,0000
25,43	16	0,0626

Źródło: badania własne

Poziom istotności pierwszej wartości chi-kwadrat wynosi $\alpha = 0,0369$, co oznacza, że hipotezę niezależności odrzuca się z ryzykiem wynoszącym 3,69%. Poziom istotności drugiej wartości wynosi $\alpha = 0,0041$, wobec czego następuje odrzucenie hipotezy niezależności z ryzykiem 0,41%. W trzeciej sytuacji nie ma podstaw do odrzucenia hipotezy niezależności, w przypadku czwartym poziom istotności wynosi $\alpha = 0,0000$ i ryzyko dla odrzucenia hipotezy jest mniejsze nawet od 0,00%, w piątym – ponownie nie odrzuca się hipotezy niezależności.

Wskazane wyżej przykłady pokazują, że jeżeli wartość numeryczna w rubryce „poziom istotności” okazuje się mniejsza niż 0,05, wówczas odrzuca się hipotezę niezależności (lub – na ogół – hipotezę zerową), a jeżeli wartość ta przewyższa 0,05 – do odrzucenia nie dochodzi. Taki raport zawiera wszystko, co jest potrzebne do zrozumienia interpretacji i do przeprowadzenia późniejszych metaanaliz. Często w artykułach w czasopiśmie naukowych zdarza się, że – z powodów ekonomicznych – wyniki testu nie zawierają tabel z danymi procentowymi. W takiej sytuacji należy dopełnić raport testu dodatkowo o liczebność próby.

Zakończenie

Statystyczne testowanie hipotez przedstawia ważne i skuteczne narzędzie w badaniach empirycznych. Różnorodne testy obejmują wszystkie możliwe sytuacje badawcze. W każdym wypadku testy te okazują się niezastąpione w badaniach na próbach podczas generalizacji wyników z prób na całe populacje. Korzystając z testów, należy jednak precyzyjnie rozumieć ich istotę i zdawać sobie sprawę z tego, jakie są ich możliwości i ograniczenia. Interpretacja wyników jest procedurą skomplikowaną, ponieważ granica między poprawnymi i niewłaściwymi wnioskami pozostaje bardzo niewyraźna dla słabo przygotowanego badacza.

Bibliografia

- Alexander, C.J., Scozzaro, M.J., Borodkin, L.J. [1989], *Statistical and empirical examination of the chi-square test for homogeneity of correlations in meta-analysis*, „Psychological Bulletin”, 106, s. 329-331.
- Babbie, E. [2004], *Badania społeczne w praktyce*, Warszawa.
- Brzeziński, J. [2003], *Metodologia badań psychologicznych*, Warszawa.
- Cohen, J. [1988], *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale.
- Ferguson, A.G., Takane Y. [1997], *Analiza statystyczna w psychologii i pedagogice*, Warszawa.
- Sobocki, M. [2000], *Metody i techniki badań pedagogicznych*, Kraków.
- Oakes, M. [1986], *Statistical inference: A commentary for the social and behavioral sciences*, New York.
- Pilch, T., Bauman, T. [2001], *Zasady badań pedagogicznych, Strategie ilościowe i jakościowe*, Warszawa.
- Rosnow, R.L., Rosenthal, R. [1989], *Statistical procedures and the justification of knowledge in psychological science*, „American Psychologist”, 44, s. 1276-1284.