

SŁAWOMIR PASIKOWSKI
ORCID: 0000-0002-0768-1596
Uniwersytet Łódzki
E-mail: slawomir.pasikowski@now.uni.lodz.pl

O MYŚLENIU STATYSTYCZNYM W PEDAGOGICE.
NA MARGINESIE ODCZYTAŃ KSIĄŻKI WIESŁAWA SZYMCZAKA
PT. *PRAKTYKA WNIOSKOWANIA STATYSTYCZNEGO*

Zaufanie do statystyki w praktyce badań pedagogicznych, a także w kształceniu akademickim pedagogów, charakteryzuje co najmniej rezerwa. Od lat wskazuje na to zawartość programów studiów pedagogicznych, w których z rzadka pojawiają się przedmioty o takich nazwach, jak statystyka, wnioskowanie statystyczne lub statystyczna analiza danych. W tym zakresie kształcenie pedagogów odstaje od kształcenia na kierunkach socjologicznych lub psychologicznych. W sporej części badań prowadzonych przez pedagogów nie sięga się po statystykę w ogóle (mimo że oferuje ona instrumenty analizy danych także jakościowych) albo korzysta się z najprostszych zestawień częstości, wskaźników struktury, ewentualnie z trudniejszych rozwiązań z przewrotnie i niechlubnie nazywanej niekiedy w pedagogice „metodologii spod znaku χ^2 ”. Raczej na marginesie wyborów pozostaje repertuar metod i rozwiązań, które nieco mniej uprzedzonemu użytkownikowi statystyki pozwalają podnosić informatywność gromadzonych danych oraz formułować wnioski wykraczające poza horyzont płaskich opowieści o „różnicach” i „zależnościach”. Z drugiej jednak strony zdarzają się ryzykowne zachowania użytkowników statystyki wynikające z naiwnej ufności pokładanej w miernikach procesów statystycznych i oprogramowaniu liczącym.

Źródeł wspomnianej niechęci oraz naiwnego zawierzania szukać można zapewne w doświadczeniach związanych z procesem kształcenia, nie tylko akademickiego, oraz w indywidualnych i grupowo podzielanych zgeneralizowanych sądach na temat statystyki. Z kolei konsekwencje tych sądów dałyby się zawrzeć pomiędzy unikaniem statystyki oraz towarzyszącej temu dezawuacji publikacji i badań, w których występują metody statystyczne, a ślełą wiarą w bezwzględną wartość wyników analiz statystycznych i w moc statystyki w zakresie odkrywania prawdy.

Między innymi z tych powodów także w przestrzeni dyskursu poświęconego metodom gromadzenia i analizy danych uwypuklają się wyraziste postaci zaciętych krytyków, jak też piewców statystyki. Wiesław Szymczak (WS) nie należy do żadnej z tych grup. W swej monografii zatytułowanej *Praktyka wnioskowania statystycznego* i wydanej nakładem Wydawnictwa Uniwersytetu Łódzkiego w 2018 roku kreśli wyraźny, także dla laika, obraz teorii statystyki, akcentując przy tym wagę uwzględniania w doborze metod założeń teoretycznych oraz warunków stosowania tych metod. WS zwraca uwagę na newralgiczne okoliczności, w których powstają nadużycia, a także na mankamenty bezmyślnego zawierzenia statystycznym metodom i informatycznym algorytmom. Krytyce poddaje konstrukty powstałe na gruncie praktyki statystyki i zauważa, że statystyka to sposób myślenia, a nie zbiór przepisów, na wzór książki kucharskiej pozwalających przygotować dobrze wyglądające porcje twierdzeń, którymi daje się żywić wyobraźnia nietroszcząca się o zasadę racji dostatecznej. Rozważania, które WS zawarł w swej monografii, uświadomią jej czytelnikom, jak łatwo jest zastosować nieodpowiednie rozwiązania, które następnie stanowią podstawę budowania rozległych i nawet powszechnie poważanych interpretacji merytorycznych, jednak niemających nic wspólnego z badanymi stanami rzeczy i ostatecznie będących fałszami. Dodałbym przy tym, że jeśli sięganie po statystykę nie jest poparte motywacją zrozumienia jej teorii, a zatem także procesów, które statystyka opisuje, to lepiej nie sięgać po nią wcale i o niej milczeć. Wypadałoby tak czynić choćby z troski o jej społeczny odbiór oraz te z zainteresowanych osób, które korzystałyby z metod statystycznych, gdyby nie głoszone stereotypy i niepowodzenia użytkowników niedoświadczonych lub rozczarowanych wymaganiami, jakie postawiło przed nimi myślenie statystyczne.

Monografia *Praktyka wnioskowania statystycznego* składa się z pięciu rozdziałów, w których Autor przybliżył i krytycznie komentuje następujące zagadnienia: teoria i praktyka wnioskowania statystycznego; możliwości, jakie w statystycznej analizie danych stwarza oprogramowanie informatyczne; moc testów statystycznych; ocena wielkości efektu; podejścia do wnioskowania statystycznego alternatywne względem klasycznego ujęcia. Poruszając się mniej więcej wzdłuż tak określonego obszaru, poświęcę uwagę wybranym zagadnieniom, które WS porusza w swojej książce, a zarazem będę się starał przybliżyć kluczowe kwestie w nadziei na zachęcenie do dyskusji o wadach i zaletach myślenia statystycznego w pedagogice. Jeśli się to nie uda, to przynajmniej do sięgania po przywołaną monografię, jak też inne podobne jej publikacje, wśród których tą, która wydaje się rezonować szczególnie, jest *Myślenie statystyczne* Walentego Ostasiewicza¹.

¹ W. Ostasiewicz, *Myślenie statystyczne*, Wolters Kluwer, Warszawa 2012.

Wnioskowanie statystyczne stanowi postać rozumowania i jako takie podlega zasadom opisywanym przez logikę. Stąd jednym z pierwszych zdań, które odczytać można w książce WS, jest zwrócenie uwagi, że jakość przesłanek we wnioskowaniu przesądza o jakości wniosku. Ma to wyraz także w znanym powiedzeniu „Garbage In, Garbage Out”. Choć powiedzenie to zawęża kwestię charakteru związku przesłanek i wniosków.

WS już w przedmowie podkreśla istnienie rozbieżności między teorią a praktyką statystyki. Statystycy, w ślad za dostrzeganiem tej rozbieżności, poddają teorię statystyczną rozwinięciom pozwalającym na zastosowanie określonych metod pomimo braku spełniania konkretnych założeń, jakich wyjściowo wymaga się od warunków używania tych metod. Te warunki determinuje sposób definiowania mierzonych zjawisk oraz sposób pozyskiwania danych, ale dzieje się to nadal w granicach twierdzeń teorii statystycznej. Oczywiście, osobną sprawą jest szkodliwy obyczaj stosowania metod statystycznych bez sprawdzania założeń znajdujących się u ich podstaw. Ramy wielu z tych założeń określa przede wszystkim teoria prawdopodobieństwa. Stanowi ona konieczną podstawę wnioskowania statystycznego. Innymi słowy, bez operowania pojęciem prawdopodobieństwa sięganie po metody rozwijane w nurcie wnioskowania statystycznego jest co najmniej beztroskie, a oczekiwanie obnażania się faktów albo prawdy na skutek zastosowania samych tylko metod stanowi najpoważniejsze z niepoważnych nieporozumień. Badane zjawisko wymaga, poza teorią, która je modeluje, również teorii określającej niepewność zdarzeń i procesów, w których to zjawisko się realizuje. Gdy wnioskowanie nie może być niezawodne, a z takim ostatecznie ma się do czynienia w naukach empirycznych, to pozostaje poleganie na wnioskach uprawdopodobnionych. Teoria prawdopodobieństwa tę niepewność pozwala kontrolować. To zaś uzasadnia korzystanie w praktyce badawczej z metod wnioskowania statystycznego, o którego teoretycznej podstawie była właśnie mowa.

Warto w tym miejscu przywołać ideę „prawdziwej wiedzy”, której miałyby rzekomo dostarczać wnioskowanie statystyczne. Ta idea jest mistyfikacją pochodzącą spoza statystyki i niekiedy naiwnie wykorzystywaną w kwestionowaniu wiarygodności statystyki. Po pierwsze, wiedza wytwarzana jest z danych o określonej jakości. Po drugie, i nieco innymi słowy niż kilka wersów wyżej, poznanie w naukach empirycznych, a do takich zaliczana jest także pedagogika, jest poznaniem w przeważającej mierze indukcyjnym. To oznacza, że kierunek większości rozumowań zwrócony jest od przypadku do reguły, więc generalizacje budowane na podstawie zbiorów przypadków mogą być co najwyżej uprawdopodobniane. Zatem walor metod statystycznych (z zastrzeżeniem, że stosowanych zgodnie z założeniami teorii, które leżą u ich podstaw) tkwi w możliwości szacowania stopnia niepewności wiedzy wytwarzanej przy użyciu tych metod, a nie w gwarantowaniu jej prawdziwości. Wnioskowanie statystyczne jest zatem wnioskowaniem reduk-

cyjnym, a więc zawodnym, choć korzystającym z praw logiki dedukcyjnej, np. prawa transpozycji, modus tollendo tollens, w celu kontrolowania niepewności². No i rzecz jasna wnioskowanie statystyczne wiąże się z podejściem stochastycznym, a nie deterministycznym, jak czasem sugerują komentatorzy przypisujący statystyce i statystykom aspiracje do wypowiedzania się o prawdzie.

W badaniach realizowanych na gruncie nauk społecznych spostrzec można wyraźną dysproporcję w korzystaniu z metod wnioskowania statystycznego. Metody weryfikacji hipotez statystycznych w stosunku do metod estymacji parametrów rozkładów cech są nieporównanie częściej użytkowane. WS zwraca uwagę na to zjawisko. Zapewne zwraca na nie uwagę również większość osób przyglądających się temu, jakie badania i w jaki sposób prowadzone są w obszarze pedagogiki, czy szerzej, w dziedzinie nauk społecznych. W przypadku badań realizowanych w pedagogice teoria estymacji mogłaby w stosunkowo prosty sposób podnieść jakość wnioskowań, a przy odpowiednim zaprojektowaniu tych badań zakres wniosków rozszerzyć. Oferuje niemało badaczom pedagogom oczekującym możliwości dokonywania oszacowań interesujących ich stanów lub procesów zachodzących w populacjach, a z powodów choćby ekonomicznych trudnych do obserwowania inaczej, jak w drodze selekcjonowania prób z tych populacji. Kwestią przy tym oczywistą dla czytelników podstawowych podręczników do statystyki jest rola próby we wnioskowaniu statystycznym, które polega na formułowaniu sądów o zbiorowości generalnej na podstawie zbiorowości częściowej. Nie może to być dowolna część zbiorowości generalnej, lecz taka, której dobór spełnia kryteria określone przez teorię statystyczną. Bez tego wnioskowanie statystyczne nie ma sensu.

W klasycznym podejściu do weryfikacji hipotez statystycznych wyróżniane są dwie teorie. Historycznie pierwsza zaproponowana została przez Ronalda Fishera, a druga przez Jerzego Sławę-Neymana i Egona Pearsona. WS obie te teorie przywołuje i przybliża kluczowe nieporozumienia zachodzące w ich aplikacji podczas praktyki wnioskowania statystycznego, a które wynikają z pomieszania pojęć tychże teorii. Jest tak w szczególności w odniesieniu do zagadnienia istotności statystycznej, które w teorii Fishera oznacza prawdopodobieństwo prawdziwości hipotezy zerowej, a w teorii Neymana-Pearsona prawdopodobieństwo odrzucenia hipotezy zerowej, gdy nie ma podstaw do jej odrzucenia. To prawdopodobieństwo określa się na podstawie testu statystycznego i nazywa *p value* lub po prostu *wartość p*, ewentualnie *p-wartość*, dla odróżnienia od prawdopodobieństwa α , o którym będzie mowa dalej. Ten tzw. testowy poziom istotności statystycznej w praktyce mylnie interpretowany bywa jako argument za słusnością hipote-

² W. Ostasiewicz, *Myślenie...*, dz. cyt.; C.R. Rao, *Statystyka i prawda*, Wydawnictwo Naukowe PWN, Warszawa 1994.

zy alternatywnej oraz w ramach pojedynczego badania, tj. warunków konkretnej próby badawczej. Co więcej, zwykle pomija się, że testowanie hipotez w modelu Fishera przewiduje interpretację prawdopodobieństwa p w serii doświadczeń. Otóż, wraz z odrzuceniem hipotezy zerowej, ta w modelu Fishera odpowiada hipotezie badawczej (dla odmiany, w modelu Neymana-Pearsona hipotezie badawczej odpowiada hipoteza alternatywna), odrzuca się testowany model zjawiska, a w jego miejsce proponuje się nowy. Na gruncie tych odmienności powstały dwa kanony testowania hipotez: testowanie przyjmująco-potwierdzające i testowanie odrzucająco-potwierdzające, których różnice również bywają niefrasobliwie pomijane. Inne dość często występujące nieporozumienie polega na tym, że hipotezy statystyczne identyfikuje się z próbą, podczas gdy zgodnie z teorią statystyczną są one sądami o populacji, które testowane są na podstawie danych z próby.

Nie dość powtarzać, że dla kondycji praktyki badawczej nieocenioną wartością ma świadomość teoretycznego podłoża metod, które w tej praktyce są wykorzystywane. Tak jest w szczególności z przywoływanymi tu metodami weryfikacji hipotez statystycznych, z których sięgający po statystykę pedagogzy chcieliby korzystać chyba najczęściej. Świadomość różnic dwóch klasycznych modeli weryfikacji hipotez, tj. modelu zaproponowanego przez Fishera oraz modelu Neymana-Pearsona, umożliwia uniknięcie błędów w interpretacji uzyskiwanych wyników. Te różnice wyraźnie wyeksponował WS w książce *Praktyka wnioskowania statystycznego*.

Osobną sprawą w kontekście metod weryfikacji hipotez statystycznych jest „kult istotności statystycznej”. Tak WS określa bezkrytyczne zawieranie wartości p . Stanowi ono poważne ograniczenie sensu rozumowań prowadzonych z użyciem tych metod. Po pierwsze, sama nazwa „istotność statystyczna” jest raczej pusta. Po drugie, jeśli przyjąć, że istotność statystyczna oznacza prawdopodobieństwo odrzucenia hipotezy zerowej podczas, gdy nie ma podstaw do jej odrzucenia (α), to należy zwrócić uwagę, że wartość tego prawdopodobieństwa jest ustalana arbitralnie przez badacza i w zależności od potrzeb oraz wiedzy o zjawisku, które jest badane. Błędem jest bowiem prowadzenie wszystkich wnioskowań na tym samym poziomie ufności (tj. z prawdopodobieństwem $1 - \alpha$). Ustalenie wartości merytorycznej wyniku tego wnioskowania znajduje się poza zasięgiem teorii statystycznej. Oczywiście pozostaje jeszcze sankcjonowany zwyczajem sposób myślenia o tym, co rezultaty analiz statystycznych mówią ich użytkownikom. Na to uwagę zwraca czytelnikowi również WS, oferując przejrzyste przykłady z wynikami analiz. WS podkreśla przy tym bezsens porównywania wartości p , czyli wspomnianych już tzw. testowych prawdopodobieństw odrzucenia „prawdziwej” hipotezy zerowej. Po pierwsze, skoro mamy więcej niż jedną wartość p , to oznacza to, że testowany był więcej niż jeden model. Zatem, w celu porównania modeli należałoby użyć innych metod oraz innych parametrów. Po drugie, i to WS

stanowczo akcentuje w przykładzie ze współczynnikami korelacji, porównanie tych statystyk wymaga postawienia hipotezy o równości parametrów, a następnie testowania tej hipotezy na założonym poziomie ufności ($1 - \alpha$). Działania inne niż te, mimo że realizowane pod szyldem wnioskowania statystycznego, to raczej niebezpieczne rytualizacje i zniekształcenia statystycznego myślenia. Warto w tym miejscu zaakcentować, że interpretacja wartości p polega na porównaniu jej z założonym poziomem istotności statystycznej (α). Zdarza się, że umyka to uwadze niektórych użytkowników teorii weryfikacji hipotez statystycznych.

Poza środowiskową lub dyscyplinarną legitymizacją działań i pojęć nieprzystających do teorii statystyki sytuację nieporozumień wzmacniać może oprogramowanie służące do analiz statystycznych. Powszechnie użytkowane przez przedstawicieli nauk społecznych programy takie, jak SPSS czy Statistica poza tym, że oferują nie zawsze identyczne zastawy narzędzi, to operują też ustawieniami początkowymi. Jeśli użytkownik nie rozumie znaczenia tych ustawień, to polegając na nich uznaje je mylnie za standard, a zarazem naraża się na zniekształcenia obrazu analizowanego zjawiska. Tak jest z wartością $\alpha = 0,05$ identyfikowaną bez racjonalnego uzasadnienia z analizami prowadzonymi na gruncie nauk społecznych. Poziom ufności, na jakim prowadzone będzie wnioskowanie przy założonym α , bo o tym w istocie mowa, wprawdzie jest arbitralnie ustanawiany, ale nie dowolnie. Uzależniony powinien być od wiedzy o badanym zjawisku i potrzeb w zakresie kontrolowania niepewności wnioskowania. Inna sprawa to założenia teoretyczne metod weryfikacji hipotez oraz estymacji parametrów. Kierowanie się zwyczajem skutkuje niekiedy domaganiem się zabezpieczenia warunków stosowania testów, które są już zabezpieczone w statystyce odpornością konkretnych rozwiązań na odstępstwa od tych założeń, własnościami rozkładów obserwowanych cech lub rozmiarem próby. Przy okazji dyskusowania kwestii zniekształceń osobnego komentarza wymaga zagadnienie transformacji zmiennych. W badaniach prowadzonych przez pedagogów albo nie jest ono brane pod uwagę, albo też błędnie kojarzone bywa właśnie ze zniekształceniami. W publikacjach WS³ i wielu innych autorów zagadnienie transformacji zmiennych, jako czynnik poprawiających jakość analiz, przedstawiane jest stosunkowo klarownie i na tyle przystępnie, że również pedagogzy sięgający po statystykę mogą się z nim skutecznie zapoznać.

Omawiając problematykę weryfikacji hipotez statystycznych, nie sposób pominąć zagadnienia mocy testu ($1 - \beta$), czyli tego, co w teorii Neymana-Pearsona związane jest z prawdopodobieństwem błędu II rodzaju (β) definiowanym jako praw-

³ W. Szymczak, *Podstawy statystyki dla psychologów. Podręcznik akademicki*, Difin, Warszawa 2018; W. Szymczak, *Praktyka wnioskowania statystycznego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2018.

dopodobieństwo nieodrzućenia hipotezy zerowej przez test statystyczny, podczas gdy zachodzi podstawa do jej odrzućenia. Po pierwsze, wiedza o tym zagadnieniu uzmysławia, iż testy istotności statystycznej (α) nic nie mówią o hipotezie alternatywnej, podobnie jak nie mówią nic o tym, czy stan rzeczy opisywany hipotezą zerową zachodzi, czy też nie. Dodać tu należy, że testy statystyczne wykorzystywane w praktyce są zwykle testami istotności statystycznej, np. test t-Studenta, test Fishera-Snedecora, test niezależności χ^2 , i jako takie nie pozwalają kontrolować błędu drugiego rodzaju. Po drugie, w praktyce wcale nierzadko moc testu jest pomijana. Wynika to z tego, że nie zawsze jej raportowania wymagają redakcje wydawnictw lub gremia recenzenckie oraz, jak pokazuje WS, znikoma jest nadzieja na ustalenie mocy testu przy obecnej ułomności narzędzi testowania hipotez w ramach teorii Neymana-Pearsona. Wprawdzie powstał konstrukt „empirycznej mocy testu”, ale jak dowodzi WS, z uwagi na charakter twierdzeń teorii testowania hipotezy zerowej oraz dokonywanie oceny tej mocy w trybie post hoc, raczej ponownie mamy do czynienia z mglistym pojęciem albo wręcz artefaktem. W innym miejscu WS⁴ zalicza tego rodzaju konstrukty do protez w rozumowaniu. Moc testu jest bowiem zdolnością testu statystycznego do wykrycia fałszywości hipotezy zerowej, gdy jest ona fałszywa. Ale, jak przybliża to WS w *Praktyce wnioskowania statystycznego*, zdolnością zdefiniowaną w kategoriach prawdopodobieństwa, że decyzja o nieodrzućeniu hipotezy zerowej będzie błędna nie częściej niż wartość $\beta \times 100$. Aby móc określić to ryzyko, konieczna jest znajomość prawdziwej wartości parametru, o którym mówi hipoteza alternatywna. Tego przecież nie wiadomo, bo jest to wiedza merytoryczna, wiedza o badanym zjawisku⁵, a zatem spoza modelu weryfikacji hipotez statystycznych. Ten ostatni służy, znów nie dość podkreślać, do testowania hipotezy zerowej, a nie alternatywnej. Tym samym moc testu pozostaje poza kontrolą testów istotności statystycznej. I nikogo to nie powinno dziwić. Wg WS dziwne natomiast jest to, że tę moc ustala się niekiedy post hoc. Do tego potrzebna jest już znajomość parametru porachowanego na danych w próbie albo założonego na podstawie wiedzy o rozkładzie cechy w populacji, następnie znajomość zastosowanego konkretnego testu istotności statystycznej, rozmiaru próby oraz przyjętej wartości poziomu istotności statystycznej. Ciekawego kontrgłosu w sprawie wątpliwej dla praktyki analitycznej wartości mocy testu udziela Andrzej Młodak⁶, zwracając uwagę na możliwość zarówno analitycznego, jak i symulacyjnego szacowania mocy testu. Nie zmienia to jednak zasadniczej wątpliwości, którą sformułował WS, że prawdopodobieństwo nieodrzućenia fałszywej hipotezy zerowej nie daje

⁴ W. Szymczak, *Praktyka wnioskowania...*, dz. cyt., s. 105.

⁵ Por. S.M. Kot, J. Jakubowski, A. Sokołowski, *Statystyka*, Difin, Warszawa 2011, s. 237.

⁶ A. Młodak, *Wiesław Szymczak. Praktyka wnioskowania statystycznego (Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2018)*, „Wiadomości statystyczne” 2019, 64, 8, s. 65–66.

się oszacować w modelu weryfikacji hipotez statystycznych bez sięgania po argumenty spoza tego modelu. Ponadto, jak stara się pokazać WS, podstawową wadą testowania hipotezy zerowej jest brak możliwości podjęcia decyzji o słuszności tej hipotezy. Otóż, gdy test istotności statystycznej wskazuje, że nie ma podstaw do jej odrzucenia, to nie jest to równoznaczne ze słusznością hipotezy i wskazaniem, aby ją przyjąć. Przy tym osobną sprawą jest uzasadnione symulowanie mocy testu na potrzeby planowania badań z możliwością jednoznacznego określenia planowanej liczebności próby. Interesujące uwagi w tym zakresie również można odnaleźć w *Praktyce wnioskowanie statystycznego*.

Do grupy miar statystycznych zdobywających w ostatnich latach sporą popularność w badaniach społecznych należą mierniki wielkości efektu. Wielkość efektu definiowana jest zwykle jako siła zależności pomiędzy zmiennymi, która pozostaje niezależna od wielkości próby⁷. To dość oczywista nieprawda. Przede wszystkim dlatego, że mierniki wielkości efektu wykorzystują wartości statystyk w próbie. Co więcej, mierniki te przyjmują tym wyższe wartości, im wyższe wartości przyjmuje statystyka testowa. Mimo zrozumiałej popularności – w końcu jeden miernik zastępuje myślenie statystyczne – ocena wielkości efektu nie może stanowić alternatywy dla wyników testowania istotności hipotezy zerowej. Rozumowanie, które składa się na to, co opisuje model weryfikacji hipotez statystycznych, mimo że niekoniecznie złożone, to jednak jest nadal na tyle skomplikowane, iż nie da się go zastąpić rachowaniem ilorazów albo kowariancji. Ponadto weryfikacja hipotez statystycznych, w przeciwieństwie do mierników efektu, dotyczy zwykle rozkładów cech w populacji. WS dobitnie obnaża pomysły tego rodzaju jak mierniki wielkości efektu także wtedy, gdy przywołuje argumentację rzeczników interesu miary wielkości efektu, którzy twierdzą, że miara ta pozwala oceniać praktyczne znaczenie zależności między zmiennymi. Wg WS takie sformułowania są nieprawdziwe i nieodpowiedzialne. Trudno się z nim nie zgodzić, gdy wie się, że o znaczeniu praktycznym wyniku analizy danych, nawet statystycznie istotnym, decyduje nie metoda tej analizy, lecz wiedza merytoryczna o badanym zjawisku. Za największą słabość pojęcia miary wielkości efektu WS uznaje brak podstaw teoretycznych. Można się domyślić, czytając jego komentarze w tej sprawie, że w dobie rosnącej konkurencji dla podejścia Neymana-Pearsona (to tylko w tym podejściu rozwinął się pomysł mierzenia wielkości efektu) ze strony takich choćby podejść, jak bayesowskie, neobayesowskie czy wiarygodnościowe, rozwinięcie podstawy teoretycznej tego pojęcia nie nastąpi. Najprawdopodobniej nie

⁷ M. Cypryańska, S. Bedyńska, *Porównywanie dwóch grup: testy t-Studenta i ich nieparametryczne odpowiedniki*, [w:] *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii*, red. S. Bedyńska, A. Brzezicka, Wydawnictwo Academica, Warszawa 2007, s. 196–197.

nastąpi również dlatego, że jak dotychczas, pojęcie to nie pojawiło się w pracach matematyków i statystyków⁸, co zresztą WS również odnotował. Presja instytucjonalna w postaci wyznaczania standardów obejmujących konieczność raportowania rozmiaru efektu bez racjonalnej teorii pozostaje tylko nieuzasadnionym wymaganiem. Mniej więcej w podobny sposób WS podsumowuje dyskusję na temat oceny wielkości efektu, dodając, że ocena ta stwarza jedynie pozory metodologicznej poprawności.

Ostatni rozdział *Praktyki wnioskowania statystycznego* WS poświęca dwóm alternatywnym podejściom do wnioskowania statystycznego. Pierwsze to podejście bayesowskie, które opiera się na pojęciu prawdopodobieństwa warunkowego i twierdzeniu o prawdopodobieństwie a posteriori (czyli szacowanemu na podstawie uzyskanych wcześniej wyników doświadczenia). Druga z zaprezentowanych alternatyw to podejście wiarygodnościowe, bazujące na funkcji wiarygodności wyznaczonej metodą największej wiarygodności, tą samą z którą z rzadka mają do czynienia pedagodzy, gdy wykorzystują regresję logistyczną lub regresję log-liniową i podczas których interpretują dobroć dopasowania modelu do danych empirycznych. Być może trudno oczekiwać zainteresowania tymi podejściami, gdy brakuje opanowania podejścia klasycznego. Niemniej jednak osoby ciekawe rozwiązań wychodzących naprzeciw słabościom modelu testowania hipotezy zerowej mogą mieć niewątpliwy pożytek z uwag WS na temat alternatywnych podejść do wnioskowania statystycznego. Uwagi te są jednak ograniczone, między innymi ze względu na objętość ostatniego rozdziału. Jednakże w rozdziale tym również zachowany zostaje, zgodnie z charakterem publikacji *Praktyka wnioskowania statystycznego*, krytyczny stosunek akcentujący zarówno zalety, jak i słabe strony obu tych podejść.

Na zakończenie warto jeszcze podkreślić, że gdy chodzi o teorię Neymana-Pearsona, najpopularniejszą wśród eksploatowanych teorii statystycznych w badaniach społecznych, to mimo wad jakie ma, stanowi ona zdecydowanie lepszy instrument testowania istotności hipotez niż próby jej zastępowania oparte na naiwnej wierze w sens każdej arytmetyki. Statystyka to przede wszystkim sposób myślenia, którego zawodność jest tym mniejsza, im mniejsza jest niefrasobliwość względem ram teoretycznych, w których to myślenie się realizuje. Biegłość w zakresie metod tego myślenia nie gwarantuje niezawodności wniosków. To nie jest możliwe. Pozwala jednak kontrolować niepewność poznania i formułować wypowiedzi dostatecznie uzasadnione.

Środowiskowo legitymizowana niefrasobliwość w zakresie wykorzystywania metod statystycznych lub w zakresie formułowania opinii na ich temat przyczynia

⁸ Por. np. I.N. Bronsztejn, K.A. Siemiendajew, G. Musiol, H. Mühlig, *Nowoczesne kompendium matematyki*, wyd. 6, tłum. A. Szczech, M. Gorzecki, Wydawnictwo Naukowe PWN, Warszawa 2015.

się do uznawania za dopuszczalne tego, co w istocie jest błędem lub oczywistym fałszem. Z drugiej zaś strony, do kwestionowania rozwiązań i wniosków, które mają uzasadnienie w ocenie warunków i w rygorze zastosowanych metod. Powstawanie takich książek, jak *Praktyka wnioskowania statystycznego* daje nadzieję, że w kontakcie z nimi czytelnik zwątpi w szkodliwe stereotypy dotyczące statystyki. Jak również, że lektura tego rodzaju publikacji uwypukli, iż myślenie statystyczne i jego instrumenty mają wprawdzie ograniczenia, lecz gdy je sobie uświadomić, to wzrasta wartość zastosowań tego myślenia oraz powstają warunki do wyważonej oceny wykorzystywania statystyki w badaniach, także pedagogicznych. Można przy tym życzyć sobie i wielu środowiskom akademickim, w których do znaczącego głosu dochodzą także preferencje estetyczne, że przed wydaniem sądu o wartości metod dla badań w jakimś obszarze, metody te wprawdzie zostaną rozczytane, rozpoznane i wypróbowane. Jeśli nawet okaże się, że ich użyteczność nie spełni oczekiwań, to zysk w postaci wiedzy pozwalającej być świadomym odbiorcą treści wytwarzanych z użyciem tych metod wynagrodzi zawiązką wysiłków zapoznania się z nimi oraz teoriami, na których te metody się opierają.

BIBLIOGRAFIA

- Bronsztejn I.N., Siemiendajew K.A., Musiol G., Mühlig H., *Nowoczesne kompendium matematyki*, wyd. 6, tłum. A. Szczech, M. Gorzecki, Wydawnictwo Naukowe PWN, Warszawa 2015.
- Cypryańska M., Bedyńska, S., *Porównywanie dwóch grup: testy t-Studenta i ich nieparametryczne odpowiedniki*, [w:] *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii*, red. S. Bedyńska, A. Brzezicka, Wydawnictwo Academica, Warszawa 2007, s. 184–207.
- Kot S.M., Jakubowski J., Sokołowski A., *Statystyka*, Difin, Warszawa 2011.
- Młodak A., *Wiesław Szymczak. Praktyka wnioskowania statystycznego (Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2018)*, „Wiadomości statystyczne” 2019, 64, 8, s. 61–68.
- Ostasiewicz W., *Myślenie statystyczne*, Wolters Kluwer, Warszawa 2012.
- Rao C.R., *Statystyka i prawda*, Wydawnictwo Naukowe PWN, Warszawa 1994.
- Szymczak W., *Podstawy statystyki dla psychologów. Podręcznik akademicki*, Difin, Warszawa 2018.
- Szymczak W., *Praktyka wnioskowania statystycznego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2018.

Author: Sławomir Pasikowski

Title: About Statistical Thinking in Pedagogics. On the Margin of the Reading of Wiesław Szymczak's book *The practice of Statistical Inference*

Keywords: statistical thinking, pedagogics, research in education

Discipline: Pedagogy

Language: Polish

Document type: Review article

Summary

The article is devoted to the issue of statistical thinking in pedagogy and research conducted in this discipline. Inspired by the readings of the book by Wiesław Szymczak, *The Practice of Statistical Inference*, it contains comments on popular associations and misunderstandings about what statistics is, what its principles and goals are, and therefore also comments on the using of statistical methods inconsistently with the underlying assumptions of statistical theory.