

dr hab. Piotr Radkiewicz
Instytut Psychologii PAN

Recenzja pracy doktorskiej mgr Lilianny Jarmakowskiej-Kostrzanowskiej po tytulem
**„Klasyczne oraz bayesowskie testowanie hipotez na przykładzie porównań dwóch grup
w badaniach psychologicznych”**

Rozprawa doktorska mgr Lilianny Jarmakowskiej-Kostrzanowskiej poświęcona jest porównaniu dwóch podejść do testowania hipotez w badaniach psychologicznych: klasycznemu i bayesowskiemu. Do postawienia takiego problemu badawczego zainspirowała Autorkę tocząca się od wielu lat dyskusja nad mankamentami tradycyjnych procedur testowania istotności hipotezy zerowej. Warto dodać, że w związku z tzw. kryzysem replikacyjnym dyskusja ta przeniosła się ostatnio z obrzeży środowiskowych debat do samego ich centrum. Oczywista stała się potrzeba radykalnego ulepszenia obowiązującego od dziesięcioleci warsztatu metodologiczno-statystycznego. Fakt, że z dużo większą niż wcześniej atencją zaczęto traktować potencjał statystyki bayesowskiej można w dużej mierze przypisać rosnącemu przekonaniu, że jednym ze sposobów sanacji powinno być jego dużo szersze niż dotychczas wykorzystanie w praktyce badawczej (a być może nawet zastąpienie nim metod tradycyjnych). Można zatem powiedzieć, że problematyka dysertacji mgr Jarmakowskiej-Kostrzanowskiej bardzo dobrze wpisuje się aktualne potrzeby dyscypliny.

Swoje rozważania o tym „która z konkurencyjnych szkół testowania hipotez jest lepsza?” Autorka rozpoczyna od sformułowania bezspornego i dla mnie poglądu, iż dla praktyki badawczej największe znaczenie ma pytanie o zgodność wniosków wypływających z podejścia frekwentystycznego (klasycznego) i bayesowskiego oraz pytanie o wielkość błędów decyzyjnych, na które naraża się stosujący je badacz. Z opisu aktualnego stanu wiedzy dowiadujemy się, że dotychczasowe badania prowadzone przez przedstawicieli nauk społecznych za pomocą symulacji komputerowych, wskazują na przynajmniej jedną, ale za to bardzo istotną przewagę podejścia bayesowskiego. Pokazują one, że w porównaniu z metodą klasyczną, podstawowe kryterium decyzyjne (tzw. czynnik Bayesa) stosowane w podejściu

bayesowskim wyraźnie obniża ryzyko popełnienia błędu drugiego rodzaju (czyli relatywnie rzadziej prowadzi do przyjęcia hipotezy zerowej wówczas, kiedy należałoby ją odrzucić).

Powyższe konkluzje odnoszą się do testu t-Studenta dla dwóch prób niezależnych. Ponadto mają one zastosowanie jedynie do tej klasy problemów badawczych, kiedy rozkład cechy w populacji jest normalny. Nie wiadomo zatem, czy taka sama relacja pomiędzy kryterialną wartością p a czynnikiem Bayesa utrzymuje się w przypadku innych testów statystycznych. Ale przede wszystkim czy utrzymuje się ona w porównaniach między dwiema grupami, gdy rozkład cechy odbiega od normalnego. I właśnie ta druga kwestia staje się dla Autorki głównym problemem badawczym i *clou* empirycznej części pracy. Autorka zauważa, że wbrew temu, co wydaje się sądzić wielu badaczy, wiedza którą dzisiaj dysponujemy, wskazuje że rozkłady cech psychologicznych stosunkowo rzadko przyjmują kształt rozkładu normalnego. Wynika z tego, że wcześniejsze badania, oparte na założeniu o rozkładzie normalnym, pozostawiają bardzo istotną lukę - niejasność co do relacji pomiędzy podejściem klasycznym i bayesowskim w sytuacji, gdy rozkład badanej cechy odbiega od normalnego. Lukę tę podejmuje się wypełnić mgr Jarmakowska-Kostrzanowska. Zastrzegając przy tym, że ponieważ nie jest możliwe, aby praca objęła wszystkie najpopularniejsze w badaniach psychologicznych testy, jako naturalny wybór - dopełniający istniejącą już wiedzę o tym narzędziu - traktuje test t-Studenta dla dwóch prób niezależnych.

Część teoretyczna pracy składa się z czterech rozdziałów. Dwa pierwsze poświęcono prezentacji teoretycznych podwalin klasycznego i bayesowskiego testowania hipotez. Rozdział trzeci przedstawia zestawienie najważniejszych pojęć i właściwości obu metod. W rozdziale czwartym Autorka przechodzi od ogólnego porównania obu podejść do porównania właściwych dla nich schematów wnioskowania, w sytuacji gdy badacz testuje wielkość różnic pomiędzy dwiema grupami. Rozdział ten kończy wyprowadzenie, na podstawie ogólnej formuły, wzoru na czynnik Bayesa dla dwóch prób badawczych.

Empiryczną część pracy otwierają rozdziały 5 i 6, w których Autorka *explicite* formułuje cel pracy i konkretne pytania badawcze, a następnie opisuje kolejne etapy zaplanowanej procedury badawczej. Rozdziały 7-9 to szczegółowy i bardzo metodyczny opis kroków analitycznych prowadzących Autorkę do odpowiedzi na postawione przez siebie pytania. Rozdział 7 przedstawia modelowanie gęstości rozkładu wybranej zmiennej psychologicznej, dopasowanie do niej adekwatnego rozkładu teoretycznego (rozkład *ex-gaussowski*) oraz dobór apriorycznych prawdopodobieństw parametrów badanej zmiennej w populacji (rozkład zmiennej ma kształt normalny). Końcowym efektem tych zabiegów jest wzór na czynnik Bayesa, który posłuży w dalszej części pracy za odpowiednik statystyki t-Studenta w podejściu

klasycznym. W rozdziale 8 otrzymujemy wyniki zestawienia testów różnic między dwiema próbami, wygenerowane dzięki symulacjom komputerowym osobno dla dziesięciu prób o różnych liczebnościach (od $N=10$ do $N=300$). Autorka przedstawia szczegółowe zestawienia wyników dziewięciu symulacji otrzymanych dzięki kombinacji trzech wariantów apriorycznego prawdopodobieństwa (informatywne zgodne z rzeczywistym kierunkiem zależności, informatywne niezgodne z rzeczywistym kierunkiem zależności oraz nieinformatywne) i trzech wariantów wielkości efektu w ujęciu D-Cohena ($D=0,2, 0,5$ i $0,8$). Możemy dzięki temu porównać wielkości błędów I i II rodzaju oraz odsetki zgodnych i niezgodnych decyzji uzyskanych w oparciu o statystykę t-Studenta i czynnik Bayesa (w podziale na dziesięć różnych wielkości próby, trzy warianty prawdopodobieństwa apriorycznego i trzy warianty wielkości efektu). Podobne zestawienie otrzymujemy w rozdziale 9 dla rozkładu tej samej zmiennej uzyskanego w rzeczywistych badaniach przeprowadzonych na dużej próbie respondentów. W tym przypadku symulacje dotyczą jedynie konkretnej, empirycznie zaobserwowanej wielkości efektu i jednej wartości prawdopodobieństwa apriorycznego (informatywne zgodne z rzeczywistym kierunkiem zależności).

Główny problem badawczy pracy został zoperacjonalizowany w postaci trzech pytań (Autorka powstrzymuje się przed postawieniem jednoznacznych hipotez) dotyczących tego: 1) jaki jest poziom błędów I i II rodzaju otrzymanych przy użyciu podejścia klasycznego i bayesowskiego; 2) jaki jest wpływ wielkości próby na zgodność między decyzjami generowanymi przez oba podejścia; i 3) jaki wpływ na zgodność decyzji ma wielkość próby. Zdaniem Autorki, z tożsamością obu podejść mielibyśmy do czynienia, gdyby wszystkie decyzje o przyjęciu i odrzuceniu były takie same. Natomiast o nadrzędności podejścia bayesowskiego moglibyśmy mówić, gdyby skutkowało popełnieniem mniejszej liczby błędów I i II rodzaju oraz gdyby czynnik Bayesa, przy tej samej wielkości efektu, potrzebował mniejszej liczebności prób do wykrycia różnic lub przy tej samej liczebności próby wykrywałby mniejsze różnice.

Jakie zatem są ostateczne konkluzje recenzowanej pracy? Porównanie obu podejść pokazuje, że zgodność decyzji osiągniętych przy ich użyciu zależy od wielkości efektu, wielkości próby i rodzaju prawdopodobieństwa apriorycznego. Dla informatywnego prawdopodobieństwa apriorycznego pełną zgodność oba podejścia osiągają przy następujących kombinacjach wielkości próby i wielkości efektu: $N=200$ gdy $D=0,8$, $N=300$ gdy $D=0,5$ i $N=600$ gdy $D=0,2$. Konkluzja pierwsza jest oczywista - decyzje uzyskiwane przy użyciu obu metod nie są tożsame. Powstaje zatem pytanie: czy wyniki przedstawione przez Autorkę faworyzują jedną z metod, wskazując na jej większą efektywność? Stosując przedstawione

powyżej kryteria, okazuje się - po pierwsze - że w podejściu klasycznym błąd pierwszego rodzaju jest niezależny od wielkości próby i utrzymuje się na stałym poziomie, podczas gdy w podejściu bayesowskim spada wraz ze wzrostem liczebności próby. Podejście bayesowskie wydaje się górować również w odniesieniu do wielkości błędu drugiego rodzaju, który jest wyraźnie mniejszy niż w podejściu klasycznym. Z tym jednak zastrzeżeniem, że sytuacja ulega diametralnej zmianie, kiedy aprioryczne prawdopodobieństwo estymowanych parametrów ma charakter nieinformatywny, oparty na niewiedzy badacza (zwłaszcza przy małej wielkości efektu błąd II rodzaju dla podejścia bayesowskiego jest zdecydowanie większy). Po drugie, okazuje się, że testowanie bayesowskie bynajmniej nie odznacza się szybszym wykrywaniem różnic pomiędzy średnimi grupowymi przy mniejszej liczebności próby (w szczególności nie posiada tej zalety, kiedy wielkość efektu nie jest duża). Po trzecie, podejście bayesowskie szybciej wykrywa różnice pomiędzy średnimi grupowymi, pod warunkiem jednak, że opiera się na prawdziwym prawdopodobieństwie apriorycznym.

Reasumując, porównywane metody – klasyczna i bayesowska – z pewnością nie są tożsame. Wydaje się, że podejście bayesowskie jest nieco bardziej efektywne, ale przesłanki do wyciągnięcia takiego wniosku nie są szczególnie silne i są obłożone szeregiem zastrzeżeń. Dodatkowo, dużym cieniem na prawomocność porównań pomiędzy obiema metodami kładzie się pewna specyficzna właściwość metody bayesowskiej - możliwość wydawania niekonkluzywnych decyzji (oznaczającej, że hipotezy badawczej nie należy ani przyjmować, ani odrzucić).

NAJWAŻNIEJSZE WALORY PRACY

1. Cieężar gatunkowy tematu

Problem badawczy, który postawiła przed sobą Autorka jest arcyciekawy i bardzo ważny. Nie dla jednej czy drugiej gałęzi psychologii, ale po prostu dla całej dyscypliny. Jak słusznie zauważa Autorka, od dłuższego czasu, z coraz bardziej prominentnych środowisk naukowo uprawiających psychologię, płyną apele o głębszą refleksję nad obowiązującym od dziesięcioleci kanonem wnioskowania statystycznego. Rosnąca popularność tzw. statystyki bayesowskiej wydaje się jednym z przejawów tego zjawiska. Co prawda piszący te słowa szczerze wątpi, by zmiana paradygmatu statystycznego mogła być istotnym remedium na bolączki takie jak te, które ujawnił kryzys replikacyjny minionej dekady, to jednak i dla niego nie ulega wątpliwości, że - choćby ze względu na paliatywną funkcję postępu - możemy spodziewać się, raczej prędzej niż później, istotnej rewizji tradycyjnych praktyk. Zakładając, że znacznie szersza niż dzisiaj implementacja podejścia bayesowskiego będzie jej częścią,

prace takie jak rozprawa doktorska mgr Jarmakowskiej-Kostrzanowskiej wydają mi się niezwykle potrzebne. Jest to nie tylko przystępnie napisany wykład o teoretycznych podstawach postulowanych zmian, ale przede wszystkim klarowny, dobrze uzasadniony empirycznie materiał poglądowy pokazujący oferowane przez nie zyski (i kłopoty).

2. Wprowadzenie w meritum problemu badawczego

Cześć teoretyczna pracy, obejmująca rozdziały 1-4, jest znakomita. Trudno mi sobie wyobrazić lepsze przedstawienie obu porównywanych modeli statystycznych i zasad, na których opiera się w nich testowanie hipotez. Ta część pracy jest napisana w tak klarowny i przystępny sposób (i bynajmniej bez nadmiernych uproszczeń), że moim zdaniem powinna zostać (być może wraz z częścią empiryczną) opublikowana i szeroko rozpropagowana w środowisku akademickim (zarówno na potrzeby dydaktyki, jak i dla własnych potrzeb badaczy). Zrozumiałą wśród psychologów abominacją do wzorów i formuł matematycznych (których i tu nie brakuje) proponuję rozbrajać stwierdzeniem, że w tym przypadku jest ich dokładnie tyle, ile być powinno (a i te, które są jakoś dziwnie nie „bola”).

3. Sformułowanie problemu badawczego i celów pracy

Uważam, że problem badawczy i cele (rozdziały 5 i 6), które postawiła sobie Autorka, są adekwatne do przedstawionych przesłanek teoretycznych i empirycznych, jak również bardzo dobrze odpowiadają na oczekiwania, które powinniśmy mieć względem rozpraw doktorskich. Zgadzam się z oceną Autorki, że jej cele zostały dobrane w sam raz dla tego typu pracy - ani zbyt wąsko, ani zbyt szeroko (natomiast w kwestii wąskości vs. szerokości późniejszej interpretacji wyników mam nieco inne zdanie). Należy również docenić to, że Autorka rezygnuje ze stawiania twardych hipotez, poprzestając na trzech kluczowych pytaniach badawczych. Jest to rezygnacja w pełni świadoma i dobrze uzasadniona, a wartość poznawcza pracy z pewnością na tym nie ucierpiała. Również kryteria mające rozstrzygać o „istotności” różnic pomiędzy podejściami wydają mi się trafnie dobrane (choć warto zwrócić uwagę, że kryteria 2 i 3 są silnie zależne, czy nawet pochodne od kryterium 1). Fakt, że finalnie nie dostarczają bardzo mocnych rekomendacji na rzecz jednego z podejść należy raczej uznać za inherentny efekt badanego problemu a nie skutek słabej mocy różnicującej kryteriów.

4. Logika procesu badawczego

Wydaje mi się, że słowo „znakomita” jest również właściwe do oceny procedury badawczej i przyświecającej jej logiki postępowania. Innymi słowy, zarówno droga do osiągnięcia zakładanego celu, jak i wybór użytych do tego narzędzi budzi moje wielkie uznanie. Muszę tu

jednak zrobić dość istotne zastrzeżenie. Moją ocenę rozdziałów 7, 8 i 9 należy traktować z pewną ostrożnością ze względu na brak własnych doświadczeń i dokonań w tego rodzaju pomysłach badawczych. Siłą rzeczy zatem analiza adekwatności wybranej drogi i instrumentarium musiała być obciążona czynnikiem subiektywizmu. Zadałem sobie po prostu ćwiczenie poznawcze polegające na próbie wymyślenia lepszej/bardziej przekonującej metody odpowiedzi na postawione przez Autorkę pytania.

UWAGI KRYTYCZNE

1. Porównanie podejścia klasycznego i bayesowskiego a niekonkluzywność czynnika Bayesa

Sądzę, że mogą się znaleźć tacy czytelnicy rozprawy mgr Jarmakowskiej-Kostrzanowskiej, u których pojawi się wątpliwość czy Autorka nie dokonuje aby przysłowiowego porównania „jabłek z gruszkami”. W mojej ocenie największy wpływ na to może mieć specyficzna dla podejścia bayesowskiego właściwość - możliwość wydawania niekonkluzywnych decyzji. Okazuje się bowiem, że czynnik Bayesa, inaczej niż testy klasyczne, które prowadzą do odrzucenia albo przyjęcia hipotezy zerowej, może generować decyzje wskazujące na brak podstaw zarówno do odrzucenia, jak i do nieodrzućania badanej hipotezy.

Autorka oczywiście zauważa ten problem, jednak zbyt łatwo przechodzi nad nim do porządku dziennego. Tymczasem, moim zdaniem, można zasadnie argumentować, że obecność kategorii „niekonkluzywność decyzji” *de facto* uniemożliwia rzetelne porównanie efektywności obu podejść. Zilustrujmy to wskaźnikami błędu II rodzaju pokazanymi w tabeli 8 dla wielkości efektu $D = 0,2$. Widzimy tam, że dla podejścia bayesowskiego odsetek błędów II rodzaju w funkcji liczebności próby ma rozkład krzywoliniowy - najpierw jest zbliżony do zera, potem wyraźnie wzrasta, by od pewnego momentu zacząć spadać. Przy największej z testowanych wielkości próby ($N = 300$) błąd II rodzaju dla podejścia klasycznego wynosi 0,31 a dla bayesowskiego 0,17. Zatem to drugie wydaje się wyraźnie lepsze. Czy na pewno? Autorka słusznie tłumaczy krzywoliniowość odsetków błędu II rodzaju tym, że początkowo czynnik Bayesa podejmuje niemal wyłącznie decyzje niekonkluzywne (których nie można wliczać do błędów II rodzaju). Potem dopiero, wraz ze wzrostem liczebności próby, zaczyna wyraźnie wzrastać liczba błędnie nieprzyjętych H_1 , by następnie zacząć spadać pod ciężarem konkluzywności informacji zawartej w dużych próbach. Wszystko tu się zgadza, ale moim zdaniem również przy największych liczebnościach możemy mieć do czynienia z porównaniem czegoś, co jest nieporównywalne (w tym przypadku 0,31 do 0,17). Bowiem, jak rozumiem, również i tutaj podejście bayesowskie nadal generuje pewną liczbę decyzji niekonkluzywnych. Wartość 0,17 nadal zawiera wyłącznie informacje o błędnie nieprzyjętych H_1 (bez decyzji

niekonkluzywnych), a wartość 0,31 nadal agreguje informacje o błędnie nieprzyjętych H_1 z informacjami o tych przypadkach, kiedy H_1 została błędnie odrzucona przez podejście klasyczne i jednocześnie uznana za niekonkluzywną przez podejście bayesowskie. Innymi słowy, jeśli nie mylę się gdzieś w swoim rozumowaniu, porównywalność błędu II rodzaju między oboma podejściami nigdy nie jest pełna (jeśli nie liczyć tak dużych liczebności próby, kiedy nie ma to już praktycznego znaczenia).

2. Rola apriorycznej znajomości parametrów badanej zmiennej

Kwestia, którą chcę omówić w tym punkcie do pewnego stopnia wiąże się z poprzednim pytaniem o porównywalność obu podejść, ale ma również pewne aspekty praktyczne, wykraczające poza bezpośrednią porównywalność. Z empirycznej części pracy dowiadujemy się, że zachowanie kluczowych wskaźników efektywności wnioskowania bayesowskiego w dużej mierze zależy od apriorycznej znajomości parametrów badanej zmiennej. Precyzyjnie rzecz ujmując, Autorka pokazuje, że istnieje ogromna różnica pomiędzy dwiema formami prawdopodobieństwa informatywnego a prawdopodobieństwem nieinformatywnym. Nie ma natomiast wielkiej różnicy pomiędzy prawdopodobieństwem informatywnym opartym na poprawnej wiedzy badacza i tym, które wynika z jego błędnej wiedzy.

Ta ostatnia konkluzja rodzi pytanie o teoretyczny i empiryczny sens różnicy między oboma prawdopodobieństwami informatywnymi. Dla mnie osobiście druga kategoria prawdopodobieństwa informatywnego pozostaje bytem bardzo enigmatycznym (zwłaszcza że daje wyniki zbliżone do pierwszej). Jakie są dokładnie kryteria odróżniające prawdopodobieństwo informatywne oparte na wiedzy niezgodnej z rzeczywistością od całkowitej niewiedzy (czyli, inaczej mówiąc, rzekomą wiedzę od ignorancji)? W przedstawionych symulacjach „informatywna ignorancja” i „informatywna wiedza” mają te same parametry, ale inne kierunki zależności. A co by było, gdyby było odwrotnie? Albo, gdyby obok błędnego kierunku zależności badacz podał również błędne wielkości parametrów? Krótko mówiąc, pozostaje z dość mglistym pojęciem o tym, czym dokładnie różni się „informatywna ignorancja” od całkowitej ignorancji (zakładam, że nie chodzi tu o wymuszenie jednostajności rozkładu).

I jeszcze wspomniany wcześniej aspekt praktyczny. Otóż mam wrażenie, że Autorka nie docenia praktycznego znaczenia konieczności podawania apriorycznych prawdopodobieństw parametrów w podejściu bayesowskim. Podejście klasyczne jest od tego wolne, co dość dobrze oddaje kontekst badawczy większości badań prowadzonych przez psychologów. Skąd brać informatywną wiedzę aprioryczną, skoro w większości przypadków jej nie ma? Robić badania

wstępne? Ale takie badania powinny być odpowiednio reprezentatywne, a więc w wielu wypadkach mogłyby nawet podwajać koszty. W tego typu sytuacjach decyzyjnych zamiana starej opcji (podejście klasyczne) na nową (podejście bayesowskie) powinna uwzględniać pełną kalkulację zysków i kosztów. Czy w kontekście konieczności posiadania „informatywnej wiedzy”, zyski (w mojej ocenie dość umiarkowane) które daje nowe podejście przynajmniej rekompensują koszty, które trzeba będzie ponieść? Nie jestem tego pewien.

3. Poziom generalizacji uzyskanych wyników

Praca nie daje jasnej odpowiedzi na pytanie o stopień generalizacji przedstawionych wyników na inne problemy badawcze. Przypomnijmy, że symulacje pokazane w części empirycznej dotyczą zgodności decyzyjnej pomiędzy statystykami klasycznymi i bayesowskimi dla różnic między dwiema średnimi grupowymi. Po zapoznaniu się rozprawą mgr Jarmakowskiej-Kostrzanowskiej pojawia się szereg pytań. Można je uszeregować na trzech poziomach, zgodnie z rosnącą możliwością ekstrapolacji wyników. Po pierwsze, czy prawidłowości, które ustaliła Autorka można uogólnić na sytuację badawczą, kiedy porównujemy więcej niż dwie średnie (test F w analizie wariancji)? Po drugie, czy możemy się spodziewać, że owe prawidłowości będą się odtwarzały dla klasycznych testów nieparametrycznych, w których badacz - ze względu na brak normalności rozkładu - testuje różnice pomiędzy parametrami innymi niż średnia i odchylenie standardowe (np. test Manna-Whitneya dla dwóch grup, test Kruskala-Wallisa dla $k > 2$ grup)? Po trzecie, czy możemy generalizować to, co ustaliła Autorka na sytuacje inne niż porównania międzygrupowe (np. związki korelacyjne)? Należy tu też dodać, że również sama Autorka w części teoretycznej daje do zrozumienia, iż nie ma podstaw by bezpośrednio generalizować na inne sytuacje badawcze wyniki porównania podejścia klasycznego i bayesowskiego dla konkretnego testu i konkretnego rozkładu. Oczywiście, pojedyncza rozprawa doktorska nie może odpowiedzieć na tak wiele pytań. Ja jednak mam na myśli nie wielokrotną replikację podejścia zastosowanego przez Autorkę do wyżej wymienionych wariantów metodologicznych, ale po prostu pogłębioną analizę „generalizowalności” uzyskanych wyników (idealnym miejscem do tego byłaby końcowa dyskusja wyników). Sądzę, że jest to coś czego będzie brakowało zdecydowanej większości czytelników recenzowanej pracy.

4. Pobieżna dyskusja wyników

W mojej ocenie końcowa dyskusja wyników zdecydowanie odbiega poziomem od reszty tekstu. Ta część powinno być czymś w rodzaju *crème de la crème*, (zwłaszcza, że praca

dostarcza sporej ilości rzeczowej „śmietanki”), a tymczasem sprawia wrażenie fragmentu napisanego w pośpiechu, w sposób bardzo minimalistyczny.

Jest tu kilka fragmentów dla mnie niejasnych. Na przykład, końcowy akapit strony 142 z nieprecyzyjnym omówieniem błędów decyzyjnych popełnianych przez oba podejścia (w pewnym momencie czytelnik traci orientację czy mowa o błędzie I czy II rodzaju); środkowy akapit strony 145, w którym pojawia się - moim zdaniem nieco naciągana - analogia z Centralnym Twierdzeniem Granicznym i nieskończonością (według mojej wiedzy, powszechnie dostępne rozkłady takich statystyk jak t-Studenta, jasno wskazują graniczne liczebności próby, przy których statystyki w próbie można traktować tak, jakby pochodziły z rozkładu normalnego); początek strony 146, gdzie mowa jest o tym, że „warto wykorzystać w analizach wyników nieistotne statystycznie, ponieważ jest szansa, że okaże się, że dane wspierają hipotezę zerową” (nie rozumiem co Autorka ma tutaj na myśli).

Więszym problemem końcowej dyskusji są jednak liczne ogólnikowe tezy, które nie znajdują potem rozwinięcia, albo jest ono dość zdawkowe. Na przykład, w ostatnim akapicie strony 143 Autorka wyróżnia teoretyczne i praktyczne aspekty odpowiedzi na pytanie o nadrzędność jednego z podejść, ale o aspektach teoretycznych czytelnik niczego się w gruncie rzeczy nie dowiaduje (z przedstawionej argumentacji może nawet wynikać, że głównym powodem by środowisko psychologiczne opowiedziało się za podejściem bayesowskim powinna być sentencja Heraklita z Efezu ‘Jedyną stałą rzeczą w życiu jest zmiana’); na stronie 144 rozwinięcia wymaga opinia, że „[...] interesujące byłoby sprawdzić to, jak silnie dane wspierają hipotezę zerową w kontekście istotności lub nieistotności wyników [...]”; w końcowym akapicie strony 146 pojawia się stwierdzenie, że „statystyka bayesowska daje większe możliwości, ale z drugiej strony więcej można w niej popsuć” (myślę, że zamiast obrazowej metafory lotniczej, czytelnik wolałby - mimo wszystko - przeczytać szersze, rzeczowe uzasadnienie tego stwierdzenia).

Ponadto sądzę, że największym mankamentem dyskusji końcowej jest brak tego, co Autorka poniekąd obiecuje czytelnikowi na początku - całościowego, rzetelnego zestawienia korzyści i trudności, z którymi wiązałoby się „przestawienie” nauk psychologicznych na podejście bayesowskie. A jeżeli rola tego podejścia miałaby polegać raczej na wsparciu metody klasycznej (co zdaje się sugerować Autorka), to i w tym przypadku byłaby wskazana jakaś bardziej szczegółowa propozycja dobrych praktyk. Mgr Jarmakowska-Kostrzanowska z pewnością posiada wystarczające kompetencje by przedstawić choćby zarys takiej „programowej odezwy”.

5. Niestaranność językowa

Muszę z przykrością zauważyć, że recenzowana praca daleka jest od standardów językowych, które powinny obowiązywać dla rozpraw doktorskich. O ile nie można mieć szczególnych zastrzeżeń co do stylu i klarowności całej narracji, to gramatyczna strona pozostawia bardzo wiele do życzenia. Mówiąc wprost - ilość błędów ortograficznych i składniowych jest wręcz zastraszająca. Szczególnie zaś irytujące są te fragmenty, gdzie Autorka w sposób mało staranny nanosi poprawki na wcześniejsze wersje tekstu, pozostawiając po nich liczne „reminiscencje”. Ekstremalnym tego przykładem są fragmenty na stronie 143, w których Autorka odwołuje się do nieistniejącego rozdziału 10 (który został włączony do rozdziałów 8 i 9), a chwilę później do końcowych tabel rozdziału 8 mając na myśli końcowe tabele podrozdziału 8.1. Nie ma to oczywiście wpływu na merytoryczną wartość pracy, ale z pewnością rzutuje na ogólne wrażenie.

Konkluzja

Uważam, że przedstawiona mi do recenzji rozprawa doktorska Pani mgr Lilianny Jarmakowskiej-Kostrzanowskiej pt. *Klasyczne oraz bayesowskie testowanie hipotez na przykładzie porównań dwóch grup w badaniach psychologicznych* spełnia wymogi stawiane przez ustawodawcę rozprawom doktorskim. Wnoszę zatem do Rady Naukowej Dyscypliny Psychologia Uniwersytetu Warszawskiego o **dopuszczenie rozprawy Pani mgr Jarmakowskiej-Kostrzanowskiej do dalszych etapów przewodu doktorskiego.**



dr hab. Piotr Radkiewicz, prof. PAN