

ANALIZA DANYCH: NOWE WYZWANIA

[*Data analysis: New challenges*]

pod redakcją gościnną Marioli Łaguny i Wiktora Razmusa

NOWE WYZWANIA W ANALIZIE DANYCH: WPROWADZENIE*

Mariola Łaguna, Wiktor Razmus

Institut Psychologii, Katolicki Uniwersytet Lubelski Jana Pawła II
Institute of Psychology, The John Paul II Catholic University of Lublin

NEW CHALLENGES IN DATA ANALYSIS: INTRODUCTION

Summary. Rapid development in statistics and methodology leads to emergence of new approaches of data analysis, which apply both to new and already existing data. Use of these new approaches allows better understanding different psychological and social phenomena. This paper is an introduction to the special part of the issue of Polish Psychological Forum dedicated to new data analytic approaches. This part contains presentation, first, method of analysis of already existing data, such as systematic literature review and meta-analysis. Next, there are presented analytical methods applied at the initial stage of data analysis: ways of treating common method variance, missing data, estimating reliability of multidimensional scales and testing measurement equivalence. Finally, there are presented approaches used at the stage of research planning and data analysis, such as dyadic research, multilevel analysis, and latent growth curve modeling applied in longitudinal research.

Key words: data analysis, research planning, methodology

Szybki postęp w dziedzinie statystyki i metodologii badań skutkuje pojawianiem się nowych metod badawczych i podejść do analizy danych (Salsburg, 2001). Wiele z tych rozwiązań jest stosowanych z powodzeniem w badaniach psychologicznych na świecie. Nie od dziś wiadomo, że psychologia staje się coraz bardziej

* Projekt został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2013/10/M/HS6/00475.

Adres do korespondencji: Mariola Łaguna, e-mail: laguna@kul.pl

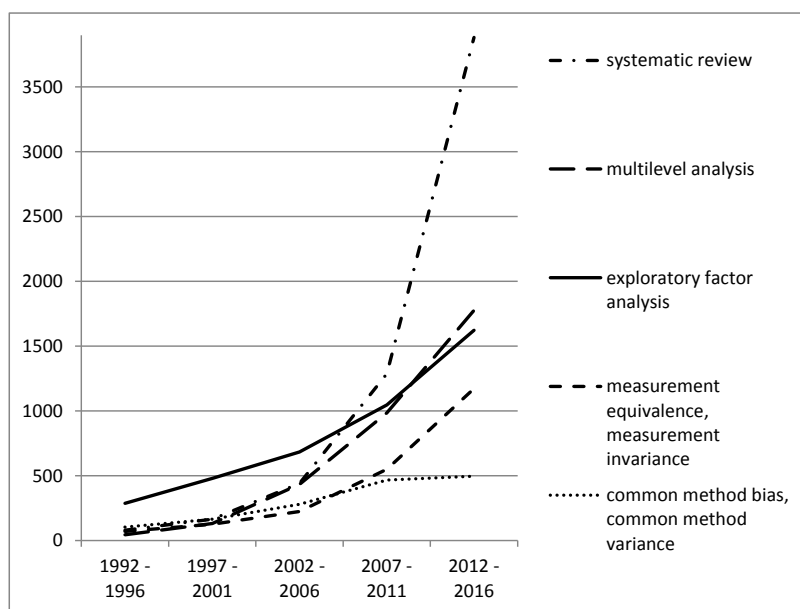
wyrafinowaną nauką, w dużej mierze dzięki rozwojowi statystyki i metodologii. Publikacje dotyczące statystyki i metodologii miały i mają duży wpływ na badania psychologiczne. Dowodem na to może być fakt, że w 1992 roku wśród 10 najczęściej cytowanych artykułów z *Psychological Bulletin* siedem odnosiło się do kwestii metodologicznych i statystycznych (Sternberg, 1992). Jeśli dziś odwiedzimy stronę tego wysoko punktowanego czasopisma (pięcioletni IF 21.971, lista A MNISW 50 punktów), to wśród pięciu najczęściej cytowanych artykułów znajdziemy cztery, które dotyczą metodologii i statystyki. Analiza cytowań wszystkich artykułów psychologicznych z bazy Web of Science wskazuje, że na pierwszym miejscu najczęściej cytowanych tekstów znalazł się artykuł Jacoba Cohena *A power primer* opisujący wiele zasad, którymi powinni kierować się psychologowie przy ocenie wyników swoich badań (Ho, Hartley, 2016). Rozwój w zakresie metodologii ma więc wyraźny oddźwięk w literaturze i konsekwencje dla sposobu prowadzenia dalszych badań empirycznych.

Postęp w dziedzinie modeli analitycznych oraz w metodologii badań pozwala coraz lepiej ujmować różnorodne zjawiska psychiczne i społeczne. Jednocześnie wymaga to od badaczy stałego poszerzania swojej wiedzy w tym zakresie i doskonalenia warsztatu metodologiczno-statystycznego. Zapoznanie się z najnowszymi podejściami nie jest zadaniem łatwym, zwłaszcza dla studentów, ponieważ niewiele jest polskojęzycznych publikacji naukowych na ten temat. W polskiej literaturze omówione zostały w przystępny sposób zarówno podstawowe metody analizy danych, takie jak np. analiza regresji, analiza wariancji czy analiza czynnikowa (np. Brzeziński, 2004a; Mackiewicz, 2007; Ferguson, Takane, 2009; Francuz, King, Minium, 2009; Bedyńska, Książek, 2012), jak i rzadziej stosowane podejścia, jak np. analiza skupień czy analiza dyskryminacyjna (np. Brzeziński, 2004b; Sosnowski, 2010a). Analiza danych rozwija się jednak stale, pojawiają się nowe podejścia zarówno do danych zgromadzonych przez badacza, czyli wyników konkretnego badania empirycznego, jak i do danych zastanych, np. w postaci wyników dotychczasowych badań. Rodzi to potrzebę ciągłego zapoznawania się z nowymi propozycjami w tym zakresie i doskonalenia warsztatu. Jednocześnie kursy akademickie, nawet na poziomie doktoranckim, rzadko zawierają nauczanie zaawansowanych i rzadziej stosowanych narzędzi analitycznych, skupiając się raczej na podstawowych metodach, jak np. analiza wariancji (Sosnowski, 2010b). Przy rosnących wymaganiach, np. ze strony redakcji czasopism, warto poszerzać repertuar dostępnych metod, które pozwalają bardziej precyzyjnie projektować badania i analizować ich wyniki, wyciągając z nich poprawne wnioski.

Wybór sposobu postępowania z danymi ma konsekwencje na różnych etapach procesu badawczego, określając w jaki sposób konkretne zjawiska będą mierzone i na ile precyzyjnie, jaki model badawczy zostanie przyjęty jako podstawa dla projektowania zbierania danych, na ile przyjęte sposoby statystycznego opracowania danych będą z nim spójne. Od poprawnego przeprowadzenia całego procesu badawczego będzie zależało czy wyniki danego badania przyczynią się do rozwoju teorii psychologicznej, pozwolą lepiej wyjaśnić dane zjawisko (Brzeziński, 2012). To bowiem „zastosowany model statystyczny formalizuje kluczowe relacje postulo-

wane przez teorię zjawiska psychologicznego” (Konarski, 2012, s. 48), a więc przyjęta teoria wyjaśniająca stanowi klucz doboru odpowiedniego podejścia do analizy danych. Chociaż kontekstowi teoretycznemu badania nie będziemy poświęcać w tym tomie zasadniczej uwagi, nie można o nim zapominać, gdyż to on stanowi zarówno punkt wyjścia badań empirycznych, jak i ich punkt dojścia.

Celem specjalnej części Polskiego Forum Psychologicznego, która została zatytułowana *Analiza danych: Nowe wyzwania*, jest przybliżenie Czytelnikom niektórych z najnowszych rozwiązań w tym zakresie. Na rosnące znaczenie tych metod wskazuje bardzo wyraźny wzrost zainteresowania nimi wśród badaczy. Korzystając z bazy PsycARTICLES przeprowadziliśmy analizę częstości odwoływania się w artykułach naukowych do kilku z omawianych w tym numerze metod analizy danych. W okresie od 1992 do 2016 roku bazę tę przeszukiwano w przedziałach pięcioletnich, a analizy dokonano w lutym 2017 roku. Jako przykładowe wybraliśmy cztery zagadnienia: błąd wspólnej metody (wyszukiwanie według dwóch haseł *common method bias* lub *common method variance*), równoważność pomiaru (wyszukiwanie według dwóch haseł *measurement equivalence* lub *measurement invariance*), analiza wielopoziomowa (*multilevel analysis*), systematyczny przegląd badań (*systematic review*). Jako punkt odniesienia wybraliśmy eksploracyjną analizę czynnikową (wyszukiwaną według dwóch haseł *exploratory factor analysis* lub EFA).



Rysunek 1. Analiza częstości występowania w publikacjach naukowych wybranych metod analizy danych

Źródło: opracowanie własne.

Jak pokazuje rysunek 1, można zaobserwować wyraźny wzrost częstości używania tych terminów, a więc także wzrost częstości używania tych metod analizy. Wzrost ten jest widoczny szczególnie w okresie ostatnich pięciu lat, po roku 2010. Różnice w tempie wzrostu częstości występowania w publikacjach różnych metod analizy, np. bardzo intensywny wzrost częstości stosowania systematycznego przeglądu badań, a słabszy testowania równoważności pomiaru, pokazują, że tendencje te nie są jedynie odzwierciedleniem generalnego wzrostu ilości publikacji naukowych na przestrzeni lat. Przybliżenie polskim psychologom i studentom psychologii nowych metod analizy danych, które stają się coraz bardziej obecne w literaturze światowej, wydaje się więc bardzo potrzebne.

Prezentowana część specjalna numeru zawiera dziewięć artykułów, które poniżej krótko przedstawimy. Kolejność artykułów została tak zaplanowana, aby odpowiadała ona strukturze procesu badawczego (Brzeziński, 2012). Zaczniemy więc od tekstów dotyczących analizy literatury przedmiotu, wskazując na możliwości usystematyzowania tego etapu badań, który poprzedza zwykle sformułowanie pytania badawczego i hipotez weryfikowanych w dalszych badaniach empirycznych. Precyzyjne zdefiniowanie zmiennej zależnej oraz wyłonienie istotnych z punktu widzenia analizowanego problemu zmiennych niezależnych wymaga także dobrania odpowiednich metod ich operacjonalizacji. Dlatego następne artykuły opisują metody ważne na etapie planowania badań oraz analizy wstępnej danych, w tym analizy właściwości zastosowanych narzędzi badawczych. Kolejne teksty dotyczą zaawansowanych metod analizy danych, rzadko uwzględnianych w nauczaniu akademickim (Sosnowski, 2010b). Ich wykorzystanie pozwala na stawianie i weryfikację nowych hipotez, która bez tego nie byłaby możliwa. Przedstawione w tym numerze artykuły nie tylko prezentują teoretyczne rozważania dotyczące różnych podejść do analizy danych, ale dostarczają również wskazówek jak praktycznie stosować omawiane metody i techniki.

Każde badanie naukowe powinno wyrastać z dostrzeżenia luk w dotychczasowym stanie wiedzy, a to z kolei jest możliwe dzięki rzetelnej analizie literatury przedmiotu. Zgłębianie literatury w interesującym badacza zakresie jest nieodzownym elementem procesu badawczego, gdyż zabezpiecza go to przed „wyważaniem otwartych drzwi”. Jest to dość oczywiste i badacze w mniejszym lub większym stopniu dbają o solidny przegląd literatury przedmiotu. Aby bardziej zobiektywizować ten proces, można skorzystać z podejścia określanego jako systematyczny przegląd badań. Wskazówek odnośnie przeprowadzenia takiego przeglądu dostarcza pierwszy z artykułów w tym tomie autorstwa Zofii Mazur i Agnieszki Orłowskiej (2018).

Kolejny artykuł autorstwa Macieja Karwowskiego (2018), opisuje metodę metaanalizy, która pozwala nie tylko na przegląd dotychczasowych badań empirycznych w konkretnym obszarze, ale także na ich analizę ilościową. W artykule omówione zostały cele stosowania metaanaliz, funkcje jakie one pełnią, przedstawiony został także proces realizowania metaanaliz, zilustrowany przykładem praktycznym. Dzięki temu czytelnik uzyska przydatne wskazówki w jaki sposób można wykorzystać tę bardzo wartościową metodę.

Analiza literatury przedmiotu kończy się zwykle sformułowaniem pytań badawczych, a czasem także sformułowaniem hipotez. W następnym kroku badacz przystępuje zwykle do zaplanowania własnych badań empirycznych. Wśród badań psychologicznych sporą część stanowią badania korelacyjne przeprowadzane z wykorzystaniem kwestionariuszy, co może jednak rodzić pewne zagrożenia. Dlatego kolejny artykuł przedstawiony w tym numerze dotyczy problematyki błędu wspólnej metody. Jest to błąd, który może powodować wiele zagrożeń, m.in. zaburzać parametry związków pomiędzy analizowanymi zmiennymi. Wiktor Razmus i Emilia Mielniczuk (2018) omawiają źródła tego błędu oraz proceduralne i statystyczne sposoby ograniczania jego wpływu na wyniki badań. O ile statystyczne sposoby kontroli błędu wspólnej metody będą przydatne po przeprowadzeniu badań, to proceduralne warto zastosować już na etapie planowania badań.

Kolejny artykuł przygotowany przez Artura Pokropka (2018) prezentuje problem braków danych, z którym ma do czynienia właściwie każdy badacz, o ile będąc w zgodzie z wytycznymi kodeksu etycznego psychologa, nie przymusi badanych do udzielenia odpowiedzi. Autor prezentuje zarówno założenia przyjmowane przy klasyfikacji braków danych, jak i metody radzenia sobie z brakami danych opierające się na różnych sposobach ich imputacji. Jako najnowsze i najbardziej skuteczne zaprezentowane zostały metoda największej wiarygodności oraz metoda wielokrotnych imputacjach.

Gdy już poradzimy sobie z brakami danych, w pierw nim przejdziemy do właściwych analiz statystycznych powinniśmy sprawdzić rzetelność użytych narzędzi badawczych. Mamy przynajmniej kilka metod szacowania rzetelności pomiaru testem. Zwyczajem stało się wykorzystywanie do tego celu współczynnika alfa Cronbacha, istnieją jednak także inne metody, takie jak współczynnik omega McDonalda. Artykuł Barbary Ciżkowicz (2018) przedstawia możliwości wykorzystania tego współczynnika w szacowaniu rzetelności wyników testu, użytecznego zwłaszcza w przypadku testów heterogenicznych.

Mając pewność, że narzędzia, których użyliśmy dostarczyły nam rzetelnych danych, możemy przystąpić do dalszych analiz statystycznych. Do najczęściej wykorzystywanych przez psychologów metod analitycznych można z dużym prawdopodobieństwem zaliczyć analizy zależności pomiędzy zmiennymi (np. współczynniki korelacji, analiza regresji, równania strukturalne) i analizy różnic międzygrupowych (np. testy różnic, ANOVA, MANOVA). Żeby móc testować różnice w natężeniu cechy pomiędzy grupami lub siłę związku pomiędzy zmiennymi w danych grupach musimy mieć pewność, że pomiar zmiennych w tych grupach jest równoważny. Czym jest pomiar równoważny, jakie mamy poziomy równoważności i jak je analizować wyjaśnia artykuł Katarzyny Lubiewskiej i Karoliny Głogowskiej (2018).

Kiedy już badacz upewni się co do odrębności analizowanych zmiennych, rzetelności i ewentualnie równoważności ich pomiaru w analizowanych grupach, poradzi sobie z brakami danych lub zastosuje inne strategie wstępnej analizy, przychodzi czas na zasadniczą analizę danych. Paradoksalnie jednak wykorzystanie

omówionych w dalszej części tego tomu podejść do analizy nie będzie możliwe, jeśli samo badanie na etapie jego konceptualizacji teoretycznej oraz planowania strategii zbierania danych nie było odpowiednio zrealizowane. Dlatego znajomość omówionych tu metod jest niezbędna już na etapie planowania badań.

Pierwsza z trzech opisanych metod analitycznych to szacowanie dopasowania w parach. Tomasz Korulczyk (2018) dokonuje konceptualnego uporządkowania perspektyw badawczych stosowanych w badaniach par. Omawia także jakie formy analizy będą najbardziej odpowiednie biorąc pod uwagę naturę danych, np. różnorodność lub niezależność elementów pary. Prezentując możliwości zastosowania tego podejścia wskazuje także ograniczenia tego paradygmatu badawczego.

Jednym z podejść analitycznych stosowanych w analizie danych pochodzących od par może być analiza wielopoziomowa. Nie jest to jednak jej jedyne zastosowanie. Możliwości wykorzystania analizy wielopoziomowej omawia Mariola Łaguna (2018), przedstawiając jej użyteczność nie tylko np. w badaniach hierarchicznych struktur, takich jak np. firma czy szkoła, ale także do analizy zjawisk intraindywidualnych. W tym przypadku to osoba jest wyższym poziomem analizy, natomiast wielokrotne pomiary jej zachowań lub ocen stanowią niższy poziom, często bardzo interesujący dla psychologów.

Dynamiczny w ostatnim okresie rozwój zainteresowania rolą czasu w funkcjonowaniu psychicznym człowieka (Roe, 2008) skutkuje pojawieniem się nowych metod analizy wzorca zmian dokonujących się w czasie. Podejście, które jest coraz szerzej stosowane w badaniach podłużnych, jakim jest modelowanie latentnych krzywych rozwojowych prezentują w swoim artykule Ludmiła Zając-Lamparska, Łukasz Warchoń i Monika Deja (2018). Omawiają w nim zarówno założenia teoretyczne tej metody, jak i wskazówki w jaki sposób można wykorzystać ją w badaniach, zobrazowane przykładami.

Mamy nadzieję, że przedstawione w tym numerze teksty pozwolą badaczom, szczególnie młodym (studentom i doktorantom), doskonalić swój warsztat metodologiczno-statystyczny, co w konsekwencji przełoży się na prowadzenie rzetelnych badań, analizowanie ich wyników zgodnie z obowiązującymi światowymi standardami i publikowanie raportów w renomowanych czasopismach. Nie można jednak zapominać, że wyniki badania empirycznego, choćby najlepiej opracowane statystycznie, są o tyle sensowne, o ile pozwalają na rozwój teorii zjawiska i tylko w świetle tej teorii mogą być adekwatnie interpretowane (Brzeziński, 2012). Dobra teoria nadaje sens uzyskanym wynikom i pozwala zrozumieć fragment rzeczywistości, który opisuje.

Literatura cytowana

- Bedyńska, S., Książek, M. (2012). *Statystyczny drogowskaz* (t. 3). Warszawa: Wydawnictwo Akademickie Sedno.
- Brzeziński, J. (2004a). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.

- Brzeziński, J. (red.) (2004b). *Metodologia badań psychologicznych: Wybór tekstów*. Warszawa: Wydawnictwo Naukowe PWN.
- Brzeziński, J. (2012). Co to znaczy, że wyniki przeprowadzonych przez psychologów badań naukowych poddawane są analizie statystycznej? *Roczniki Psychologiczne*, 15 (3), 7-40.
- Ciżkowicz, B. (2018). Omega McDonalda jako alternatywa dla alfa Cronbacha w szacowaniu rzetelności testu. *Polskie Forum Psychologiczne*, 23 (2), 311-329, doi: 10.14656/PFP20180206
- Ferguson, G.A., Takane, Y. (2009). *Analiza statystyczna w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.
- Francuz, P., Mackiewicz, R. (2007). *Liczy nie wiedzą, skąd pochodzą. Przewodnik po metodologii i statystyce nie tylko dla psychologów*. Lublin: Wydawnictwo KUL.
- Ho, Y.S., Hartley, J. (2016). Classic articles in psychology in the Science Citation Index Expanded: A bibliometric analysis. *British Journal of Psychology*, 107, 768-780.
- Karwowski, M. (2018). Metaanaliza: O ilościowym syntetyzowaniu ustaleń empirycznych. *Polskie Forum Psychologiczne*, 23 (2), 252-276, doi: 10.14656/PFP20180203
- King, B.M., Minium, E.W. (2009). *Statystyka dla psychologów i pedagogów*. Warszawa: Wydawnictwo Naukowe PWN.
- Konarski, R. (2012). Interpretacja przyczynowa wyników analizy statystycznej. *Roczniki Psychologiczne*, 15 (3), 47-50.
- Korulczyk, T. (2018). Szacowanie dopasowania w parach: Perspektywy i ograniczenia badawcze. *Polskie Forum Psychologiczne*, 23 (2), 357-376, doi: 10.14656/PFP20180208
- Lubiewska, K., Głogowska, K. (2018). Zastosowanie analizy równoważności pomiarowej w badaniach psychologicznych. *Polskie Forum Psychologiczne*, 23 (2), 330-356, doi: 10.14656/PFP20180207
- Łaguna, M. (2018). Wprowadzenie do wielopoziomowej analizy danych. *Polskie Forum Psychologiczne*, 23 (2), 377-394, doi: 10.14656/PFP20180209
- Mazur, Z., Orłowska, A. (2018). Jak zaplanować i przeprowadzić systematyczny przegląd literatury. *Polskie Forum Psychologiczne*, 23 (2), 235-251, doi: 10.14656/PFP20180202
- Pokropek, A. (2018). Wybrane statystyczne metody radzenia sobie z brakami danych. *Polskie Forum Psychologiczne*, 23 (2), 291-310, doi: 10.14656/PFP20180205
- Razmus, W., Mielniczuk, E. (2018). Błąd wspólnej metody w badaniach kwestionariuszowych. *Polskie Forum Psychologiczne*, 23 (2), 277-290, doi: 10.14656/PFP20180204
- Roe, R.A. (2008). Time in applied psychology: The study of "what happens" rather than "what is." *European Psychologist*, 13 (1), 37-52, doi: 10.1027/1016-9040.13.1.37
- Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York, NY: William Freedman.
- Sosnowski, T. (red.) (2010a). Numer specjalny: Zaawansowane metody statystyczne. *Psychologia Społeczna*, 5 (2-3).

- Sosnowski, T. (red.) (2010b). Zaawansowane metody statystyczne. Wprowadzenie. *Psychologia Społeczna*, 5 (2-3), 113-116.
- Sternberg, R.J. (1992). Psychological Bulletin's top 10 "hit parade". *Psychological Bulletin*, 112, 387-388, doi: 10.1037/0033-2909.112.3.387
- Zajac-Lamparska, L., Warchoń, Ł., Deja, M. (2018). Analiza danych podłużnych: Modelowanie latentnych krzywych rozwojowych. *Polskie Forum Psychologiczne*, 23 (2), 395-412, doi: 10.14656/PFP20180210

Streszczenie. Szybki postęp w dziedzinie statystyki i metodologii badań skutkuje pojawianiem się nowych podejść do analizy danych, zarówno nowych, jak i zastanych. Ich wykorzystanie pozwala coraz lepiej ujmować różnorodne zjawiska psychiczne i społeczne. Artykuł stanowi wprowadzenie do specjalnej części numeru Polskiego Forum Psychologicznego poświęconej nowym podejściom do analizy danych. Zaprezentowane w nim zostały najpierw metody analizy wyników wcześniejszych badań, takie jak przegląd systematyczny i metaanaliza. Następnie omówione zostały sposoby postępowania z danymi na wstępnym etapie ich analizy: radzenie sobie z błędem wspólnej metody, z brakami danych, szacowanie rzetelności metod wielowymiarowych, a także ocena równoważności pomiaru. Przedstawione zostały także podejścia wykorzystywane w planowaniu badań i analizie ich wyników, takie jak analiza diad, analiza wielopoziomowa oraz modelowanie latentnych krzywych rozwojowych w badaniach podłużnych.

Słowa kluczowe: analiza danych, planowanie badań, metodologia

Data wpłynięcia: 20.10.2017

Data wpłynięcia po poprawkach: 5.03.2018

Data zatwierdzenia tekstu do druku: 31.03.2018