

Uniwersytet Pedagogiczny w Krakowie.
Centrum Badań Funkcji Neurokognitywnych i Samoregulacji

ARKADIUSZ PRAJZNER

ORCID: 0000-0002-8681-9636
arkadiusz.prajzner@op.pl

Wybrane wskaźniki wielkości efektu w badaniach psychologicznych

Selected Indicators of Effect Size in Psychological Research

PROPOZYCJA CYTOWANIA: Prajzner, A. (2022). Wybrane wskaźniki wielkości efektu w badaniach psychologicznych. *Annales Universitatis Mariae Curie-Skłodowska. Sectio J, Paedagogia-Psychologia*, 35(4), 139–157. DOI: 10.17951/j.2022.35.4.139-157

ABSTRAKT

Odzwierciedlając zmieniającą się praktykę statystyczną w badaniach psychologicznych, w której dominuje testowanie hipotez zerowych z wykorzystaniem decyzji o poziomie istotności wyników, wskazano zalecenia dotyczące raportowania w pracach wielkości efektu. W opracowaniu przedstawiono pojęcie wielkości efektu oraz wskazano miejsce, jakie zajmuje w analizie danych w odniesieniu do istotności wyników. Celem pracy jest opisanie wybranych wskaźników wielkości efektu, a także wskazanie potrzeby zastosowania i poprawnej ich prezentacji i interpretacji w raporcie analiz prac empirycznych z zakresu nauk społecznych. Biorąc pod uwagę ograniczenia podejścia statystycznej analizy danych opartej jedynie na poziomie istotności, w opracowaniu zaprezentowano możliwości umieszczania w analizach danych wskaźnika o większym praktycznym zastosowaniu, jakim jest wielkość efektu. Wykorzystując najbardziej popularne metody analityczne, takie jak testy t Studenta, jednoczynnikowe analizy wariancji w schematach między- i wewnątrzgrupowych, a także analizy testem Wilcozona, U Manna-Whitneya, H Kruskala-Wallisa, testem Friedmana oraz uwzględniając analizy dla danych jakościowych, zaprezentowano dobrane do planów badawczych wskaźniki wielkości efektu. Ponadto opisano wykorzystanie, sposób obliczania oraz interpretację wybranych wskaźników wielkości efektu, jakimi są wskaźniki: d Cohena, g Hedgesa, delta, r_g Glas-sa, korelacja par dopasowanych r_c , eta-kwadrat, omega-kwadrat oraz epsilon-kwadrat, W Kendalla oraz f , V Cramera czy iloraz szans i ryzyko względne. Prezentację wskaźników wielkości efektu zestawiono z odpowiadającymi im planami badawczymi i rodzajem zebranych danych.

Słowa kluczowe: wielkość efektu; hipoteza zerowa; analiza statystyczna; badania psychologiczne

WPROWADZENIE

Już za sprawą W. Wundta w 1879 r. psychologia skutecznie przekształciła się w naukę empiryczną. Od tego czasu stale ulepszane są metody samego pomiaru, co wyraża się w bardziej dokładnych testach, lepiej skonstruowanych kwestionariuszach czy trafniej dobieranych wskaźnikach. Rozwojowi podlegają także sposoby analizy uzyskanych wyników, ich prezentacji i interpretacji. Klasyczny model analizy danych i poszukiwania odpowiedzi na postawione pytania badawcze wykorzystuje testowanie istotności hipotezy zerowej. Jest to najczęściej wykorzystywana metoda analizy danych. Ma jednak pewne ograniczenia, których badacze są coraz bardziej świadomi. Próbą uzupełnienia możliwości wyciągania praktycznych wniosków z badań empirycznych są wskaźniki wielkości efektu.

Mimo powszechnej znajomości wykorzystania wielkości efektu w swoich badaniach Sun, Pan i Wang (2010) wykazali, że liczba artykułów z obszaru psychologii, gdzie raportuje się te wartości, wynosi 49%, a liczba artykułów, gdzie poddaje się je interpretacji, to 57%. Inni badacze wskazali, że nawet 42% badań psychologicznych pozostaje bez podanej wielkości efektu (Fritz, Morris, Richler, 2012). Co więcej, w wielu przypadkach wielkość efektu jest całkowicie pomijana, szczególnie przy braku istotności analizowanych testów statystycznych. W tak niejednoznacznej i wielopłaszczyznowej nauce, jaką jest psychologia, zredukowano stawiane wnioski do zero-jedynkowej decyzji, opartej jedynie o wartość poziomu istotności. Jak zaznacza Kirk (2001), liczba prac opublikowanych w „Journal of Applied Psychology”, zawierających wskaźniki wielkości efektu, wzrosła z 12% do aż 77% po publikacji w 1995 r. standardów edytorskich. Do tej znaczącej poprawy, jak zauważył Kirk, doprowadził wskaźnik R^2 wykorzystywany w analizie regresji. Ten wskaźnik, automatycznie wskazywany w podstawowych pakietach statystycznych, stanowił aż 60% wszystkich prac, gdzie prezentowane były wielkości efektu. Pozostałe wskaźniki, takie jak omega-kwadrat czy f^2 Cohena, których wykorzystanie jest bardziej poprawne np. w ocenie modelu wariancji, nie przekraczały 0,5%. Jak zaznacza Kirk (2001), przygotowane automatycznie przez programy analityczne wielkości efektu nie zawsze są dobrze dopasowane do planów badawczych. Definicja, wykorzystanie i interpretacja, a także rola, jaką odgrywa wielkość efektu obok testowania hipotezy zerowej w badaniach psychologicznych, zostały omówione w niniejszym opracowaniu.

TESTOWANIE HIPOTEZY ZEROWEJ W BADANIACH PSYCHOLOGICZNYCH

1. Podejście metody NHST

W tradycyjnym podejściu istnieje system dwóch sprzecznych hipotez, których celem jest tworzenie ram wnioskowania o populacji na podstawie próby (Field, 2013; Sun i in., 2010). Hipoteza zerowa (H_0), będąca faktycznym obiektem testowania statystycznego, oraz alternatywna (H_A), stanowiąca o przewidywanym istnieniu zależności między zmiennymi, są w tym podejściu metodą odpowiedzi na postawione pytania badawcze. Metoda NHST (*Null Hypothesis Statistical Testing*), oznaczająca statystyczne testowanie hipotezy zerowej, jak opisują King i Minium (2020), jest skoncentrowana na ocenie prawdopodobieństwa wystąpienia zdarzenia opisanego w hipotezie zerowej. Osiągnięta w analizie wartość istotności odnieszona jest do kryterium decyzyjnego, jaki stanowi przyjęty poziom Alfa, który najczęściej wynosi 0,05 lub 0,01. Uzyskany wynik istotności, będący równy lub niższy od wybranej wartości Alfa, prowadzi do wniosku, że wystąpienie zdarzenia opisanego w hipotezie zerowej jest mało prawdopodobne, zatem hipoteza zerowa nie może być prawdziwa. Skutkuje to jej odrzuceniem i przyjęciem jako wniosek o populacji hipotezy alternatywnej. Na podstawie oceny prawdopodobieństwa wystąpienia sytuacji określonej w hipotezie zerowej stawia się wnioski o zależnościach między zmiennymi w populacji, których istnienie definiuje hipoteza alternatywna.

2. Ograniczenia NHST

Bezrefleksyjne stawianie wniosków w odniesieniu jedynie do wyniku istotności może posiadać ograniczoną wartość naukową (Brzeziński, 2021). Odpowiednia manipulacja liczbą obserwacji sprawi, że nawet najmniejsza różnica, bez praktycznej wartości, będzie mogła być uznana za istotną. Rozmiar próby ma duże znaczenie dla wyniku poziomu istotności, a zatem także dla odpowiedzi na postawione pytanie badawcze (Field, 2013; King, Minium, 2020; Sun i in., 2010). Liczba obserwacji ma bezpośrednie znaczenie dla wskazania liczby stopni swobody, które razem z wyliczonym wynikiem testu statystycznego wskazują jego poziom istotności. Zwiększenie liczby stopni swobody (poprzez zwiększenie liczby obserwacji) zmniejsza wartość krytyczną wymaganą do uznania wyniku testu za istotny, co jako problem zbyt dużej próby wskazał Hays (1973, za: Brzeziński, 2021). Sugeruje się także, że podejście NHST nie zawsze jest odpowiedzią na postawione pytanie badawcze z dwóch powodów: (1) testowaniu podlega zupełnie przeciwne zdarzenie opisane w hipotezie zerowej niż to, które faktycznie interesuje badacza oraz (2) hipoteza zerowa w jakimś stopniu jest zawsze fałszywa,

ponieważ dwie średnie nigdy nie są identyczne (Ferguson, 2009; Field, 2013; Grissom, Kim, 2011).

Zero-jedynkowa natura podejścia NHST jest z pewnością przydatna dla inżynierów czy fizyków, lecz w naukach społecznych może być opisywana jako niepraktyczna i nieefektywna (Ferguson, 2009). Ponadto nie wskazuje badaczom tego, co często chcieliby wiedzieć, a także utrudnia wyciągnięcie praktycznych wniosków o badanych zależnościach (Kirk, 2001; Sun i in., 2010) i nie pozwala na akumulację wiedzy w postaci metaanaliz (Bosco, Aguinis, Singh, Pierce, 2015; Grissom, Kim, 2011). Proponowanym od dawna kierunkiem w analizie wyników są wskaźniki wielkości efektu.

WIELKOŚĆ EFEKTU

1. Definicja i wykorzystanie

Opisując pojęcie wielkości efektu, Pek i Flora (2018) wskazują, że można go tłumaczyć jako pomiar siły czy wielkości, z jaką predyktor oddziałuje na zmienną wyjściową. Field (2013) dodaje, że wielkość efektu jest obiektywnym i standaryzowanym (zazwyczaj) pomiarem obserwowanego oddziaływania. Wskaźnik wielkości efektu jest uwolnioną od jednostki pomiaru próbą oszacowania stopnia, w jakim zależność między zmiennymi jest obecna w populacji (Cohen, 1998, za: Brzeziński, 2021). *Effect size*, oznaczający wielkość efektu, można uznać za miarę siły związku między zmiennymi (Bosco i in., 2015; Kühberger, Fritz, Scherndl, 2014). Może być także traktowany jako wielkość różnic między grupami czy pomiarami (Coe, 2002; Sullivan, Feinn, 2012). Wielkością efektu można też nazwać każdą statystykę, która pozwala na wskazanie stopnia, w jakim populacja różni się od oczekiwań określonych w hipotezie zerowej. Istnieje wiele wskaźników, które mogą być wykorzystane w ocenie wielkości efektu, ale w badaniach można zauważyć pewną grupę najczęściej wykorzystywanych miar (Henson, 2016) i – jak zauważa Sun i in. (2010) – każde z nich mają wady i zalety w określonych warunkach czy planach badawczych.

Wskazanie w analizie danych wielkości efektu było zalecane już w czwartej edycji standardów edytorskich (American Psychological Association [APA], 1994) i zostało jeszcze mocniej zaznaczone w siódmej edycji (APA, 2020, s. 89): „Aby czytelnicy mogli docenić wielkość lub wagę wyników badań, zaleca się uwzględnienie pewnej miary wielkości efektu w sekcji wyników. (...) Ogólną zasadą, której należy przestrzegać, jest dostarczenie czytelnikom wystarczającej ilości informacji, aby ocenić wielkość obserwowanego efektu”. Brzeziński (2021, s. 231) również zaznacza, że „nie można wypuścić w świat pracy, w której nie będą podane wartości wielkości efektu”. Przedstawienie omawianych miar jest konieczną informacją w pracy empirycznej z kilku powodów. W pierwszej

kolejności wielkość efektu jest praktyczną wskazówką o wynikach badania i pozwala na lepsze odniesienie wyników próby na zależności zmiennych w populacji (Fritz i in., 2012; Vascha-Haase, Thompson, 2004). Dodatkowo wskaźnik ten pozwala na wykorzystanie doniesień z badania w metaanalizie (Pek, Flora, 2018; Sun i in., 2010), czego powodem może być uznanie wielkości efektu za wskaźnik wolny od liczby obserwacji (Ferguson, 2009). Wartości testów statystycznych oraz wartości poziomu istotności zawsze będą uwzględniały liczbę obserwacji, wykorzystane narzędzia i skale pomiaru, co ogranicza możliwość porównania wyników z różnych badań. Ograniczenia te, nieuwzględnione w podejściu NHST, sprawiają (szczególnie przy nielosowym doborze), że tworzenie wniosków jedynie na podstawie poziomu istotności może się stać zadaniem trywialnym (Ferguson, 2009).

2. Poziom istotności a wielkość efektu i przedziały ufności

Bardzo niski wynik poziomu istotności nie powinien stanowić podstawy do bezrefleksyjnych wniosków o istnieniu akceptowanego efektu w populacji (Sun i in., 2010). Jak zaznacza Brzeziński (2021), sam poziom istotności nie służy do oceny rzeczywistego oddziaływania między zmiennymi, tylko daje badaczom informacje o prawdopodobieństwie popełnienia błędu pierwszego rodzaju. Wartości poziomu istotności oraz wielkości efektu uzupełniają się i można o nich myśleć jak o stronach tej samej monety, lecz wnioski płynące z interpretacji jednego i drugiego wyniku nie są tożsame. Analizując wspólnie poziom istotności oraz wielkość efektu, można wskazać cztery możliwe do uzyskania w badaniach własnych sytuacje (Sun i in., 2010):

- a) istotność statystyczna + silny, akceptowalny efekt,
- b) brak istotności statystycznej + słaby, nieakceptowalny efekt,
- c) istotność statystyczna + słaby, nieakceptowalny efekt,
- d) brak istotności statystycznej + silny, akceptowalny efekt.

O ile poziom istotności może być rozpatrywany binarnie (istotny vs nieistotny wynik), o tyle wielkość efektu jest stopniowalna, zatem uzyskanie zerowego efektu jest rzadkie, stąd określenie „efekt nieakceptowalny”, sugerujący niskie i bardzo niskie jego wartości. Niewiele trudności powinna sprawić interpretacja sytuacji *a* i *b*, kiedy wartość poziomu istotności oraz wielkości efektu są spójne. Z kolei sytuacja *c* może stanowić najczęściej spotykany błąd wnioskowania oparty na wyniku poziomu istotności, kiedy osiągając bardzo niski jego wynik, wyprowadza się konkluzje na temat sposobu oddziaływania zmiennych w populacji, nie zwracając uwagi na fakt małego lub bardzo małego efektu (Brzeziński, 2021; Sun i in., 2010). Sytuacja *d* to przykład, kiedy wyraźny efekt nie może zostać zgeneralizowany i nie może prowadzić do wniosku o populacji. Ostatnie sytuacje (*c* i *d*) oraz różnice między wartością poziomu istotności a wielkością efektu mogą

być uznawane za zagrożenie w procesie walidacji badania, którego dotyczą (Sun i in., 2010). W promocji dobrych praktyk statystycznych Kirk (2001) zaznacza, że najlepszym podejściem do wskazywania doniesień empirycznych jest połączenie wniosków o wielkości efektu i o jego istotności. Oznacza to także sytuacje, kiedy otrzymana wielkość efektu jest niska lub bardzo niska. Praktyka niewykazywania niskich wartości wielkości efektu może sprawić, że do metaanaliz trafią jedynie badania i wyniki osiągające wyższe wartości. Może to doprowadzić do przeszacowania realnego związku, jaki występuje między zmiennymi, kiedy ocenie poddamy tylko wyniki i badania potwierdzające ich silną korelację.

Możliwość pełnego wykorzystania wyników badania w metaanalizach wymaga, obok wartości otrzymanej wielkości efektu, podania przedziałów ufności CI (*Confidential Intervals*; zob. Cohen, 1990, za: Brzeziński 2021). W badaniach można spotkać najczęściej 85-, 95- i 99-procentowe przedziały ufności. Ich dolna wartość LL (*Lower Limit*) oraz górna wartość UL (*Upper Limit*) stanowią przedział wyników, w którym zawiera się wynik otrzymany wielkości efektu badania. Zwiększając prawdopodobieństwo przedziałów np. z 85% do 95%, wzrośnie również rozpiętość uzyskanych wartości i jednocześnie obniży się ich precyzja. Odwrotnie będzie, gdy obniżymy prawdopodobieństwo – wówczas przedziały staną się węższe, bardziej precyzyjne (Brzeziński, 2021). Interpretacja CI stanowi czasami kłopot, ponieważ jej wartości nie oznaczają, że z danym prawdopodobieństwem przedział ten zawiera estymowany parametr populacji. Przykładowo 95-procentowy przedział ufności mówi, że powtarzając wiele razy te same pomiary na tej samej populacji, 95% tych badań będzie zawierać wartość, jaka opisuje tę populację (Field, 2013). Mając na uwadze konieczność wskazywania w analizie danych wskaźnika wielkości efektu, w dalszej części opracowania zaprezentowano wybrane wzory wraz z omówieniem i propozycją ich interpretacji.

WYBRANE WSKAŹNIKI WIELKOŚCI EFEKTU

Istnieje wiele wskaźników, które mogą być wykorzystywane w analizie danych (Kirk, 2001). Ich liczba uniemożliwia przedstawienie wszystkich możliwości, dlatego w opracowaniu zdecydowano się opisać wybrane wskaźniki, dopasowane do najczęściej wykorzystywanych analiz. Mnogość powstałych wskaźników wielkości efektu poddaje się próbie ich kategoryzacji. Można wyróżnić podział uwzględniający wskaźniki w ocenie różnic, siły związku oraz szans. W pierwszej grupie wskaźników znajdują się wskaźniki: d Cohena, g Hedgesa, Δ Glassa (Ferguson, 2009), a także transformowane wskaźniki: $\text{probit } d'$, $\text{logit } d'$, h i q Cohena (Rosnow, Rosenthal, 2003). Druga grupa wskaźników w ocenie związku wykorzystuje całą rodzinę współczynników r oraz m.in. takie jak: β , τ , φ i ρ . Do tej grupy wskaźników można zaliczyć również współczynniki kwadratowe: r^2 , R^2 , η^2 , poprawione- R^2 , ω^2 czy ε^2 (Ferguson, 2009;

Vacha-Haase, Thompson, 2004). Ostatnia wskazana grupa w analizie zmiennych nominalnych w ocenie szans i ryzyka wykorzystuje ryzyko względne i iloraz szans (Rosnow, Rosenthal, 2003).

W opracowaniu zdecydowano się zaprezentować wskaźniki wielkości efektu w odniesieniu do testów statystycznych i przykładowych planów badawczych. Prezentację podzielono na trzy obszary najczęściej wykorzystywanych testów parametrycznych, nieparametrycznych oraz analiz dla zmiennych jakościowych, uwzględniając przy tym różne plany badawcze.

1. Wielkość efektu testów parametrycznych

Często spotykanym planem badawczym jest ocena dwóch niezależnych grup. Porównując je w analizie testem *t* Studenta dla danych niezależnych, do oceny wielkości efektu zaprezentowany został przez Cohena (1988) współczynnik *d*. Wartość wielkości efektu Cohena (*d*) wyraża się wzorem (Brzeziński, 2021, s. 224):

$$d = \frac{M_1 - M_2}{\sigma} \quad (1),$$

gdzie: *M* – średnia w grupie, *σ* – odchylenie standardowe w populacji. Problemem z wykorzystaniem powyższego wzoru jest odchylenie standardowe dla populacji, które zazwyczaj jest nieznane. Trzeba dokonać estymacji tej wartości na podstawie danych z próby, co zostało zaproponowane jako współczynnik Hedgesa (*g*), opisany wzorem (King, Minium, 2020, s. 380):

$$g = \frac{M_1 - M_2}{\sqrt{\frac{SS_1 + SS_2}{(n_1 - 1) + (n_2 - 1)}}} \quad (2),$$

lub

$$g = \frac{(M_1 - M_2) \left(1 - \frac{3}{4df - 1}\right)}{\sqrt{\frac{SS_1 + SS_2}{(n_1 - 1) + (n_2 - 1)}}} \quad (3),$$

gdzie: *M* – średnia w grupie, *SS* – suma kwadratów odchyleń od średniej w grupie, *n* – liczba obserwacji w grupie, *df* – stopnie swobody. Między prezentowanym wzorem *d* Cohena oraz *g* Hedgesa King i Minium (2020) stawiają znak równości. Wskazana poprawka na liczebność próby we wzorze (3) ma szczególne zastosowanie w przypadku małych grup, a jej znaczenie maleje wraz ze wzrostem wielkości *N* (Brzeziński, 2021). Prowadzi to do wniosku, że im większa liczebność próby, tym bardziej maleje różnica między wielkością efektu *g* obliczoną ze

wzoru (2) lub (3). Inną możliwością dla rekompensacji braku znajomości odchylenia standardowego w populacji jest wykorzystanie wielkości efektu Glassa (Δ), opisanego wzorem (Fritz i in., 2012, s. 8):

$$\Delta = \frac{M_1 - M_2}{SD_{kontrolna}} \quad (4),$$

gdzie: M – średnia w grupie, $SD_{kontrolna}$ – odchylenie standardowe w grupie kontrolnej. W badaniach klinicznych czy eksperymentalnych odchylenie standardowe grupy kontrolnej może być lepszym estymatorem wartości populacji niż odniesienie tego wyniku do zmanipulowanych przez badanie wartości eksperymentalnych (Field, 2013). Problem z wielkością efektu Glassa powstaje, gdy nie mamy w planie badawczym grupy kontrolnej. W analizie porównawczej kobiet i mężczyzn, osób starszych i młodszych ciężko wskazać grupę kontrolną, dlatego współczynnik Glassa będzie miał specyficzne i wąskie zastosowanie. Alternatywny i prostszy w wykorzystaniu wzór wskaźnika d Cohena może być wyrażony wzorami (Fritz i in., 2012, s. 9):

$$d = \frac{2t}{\sqrt{n-2}} \quad (5),$$

lub

$$d = \frac{2t}{\sqrt{df}} \quad (6),$$

gdzie: t – wartość statystyki testowej, n – liczba wszystkich obserwacji, df – stopnie swobody. Stosowanie tych wzorów może być przydatne głównie w sytuacji, kiedy porównywane grupy są równoliczne (Rosenthal, Rosnow, Rubin, 2000). Wymagane wartości uzyskamy podczas analizy testem t Studenta dla danych niezależnych, wykorzystując podstawowe pakiety analityczne, takie jak SPSS, Statistica czy Excel. Sytuacja badawcza, w której grupy nie są równoliczne, powinna skłonić do rozważenia w ocenie wielkości efektu wskaźnika Hedgesa.

Porównanie dwóch pomiarów powiązanych testem t Studenta dla danych zależnych w ocenie wielkości efektu także wykorzystuje współczynnik d Cohena. Z powodu korelacji między pomiarami wzory zaprezentowane wyżej będą błędnie przybliżały odchylenie standardowe dla populacji. Jego estymacje można wyrazić w postaci współczynnika Hedgesa (g), opisanego wzorem (King, Minium, 2020, s. 409):

$$g = \frac{M_1 - M_2}{\sqrt{SD_1^2 + SD_2^2 - 2r_{1,2}SD_1SD_2}} \quad (7),$$

gdzie: M – średnia w grupie, SD – odchylenie standardowe w grupie, r – wartość korelacji Pearsona. Szacowanie wartości odchylenia dla par powiązanych wyraża się innym wzorem niż prezentowane wcześniej dla grup niezależnych i jest to szacowanie uciążliwe w praktyce. Wzór ten może jednak zostać uproszczony do postaci (King, Minium, 2020, s. 409):

$$g = \frac{\bar{D}}{\sqrt{\frac{SS_D}{n-1}}} \quad (8),$$

gdzie: \bar{D} – różnica średnich pierwszego i drugiego pomiaru ($M_1 - M_2$), SS_D – suma kwadratów odchylenia wyników różnicy między pomiarami od średniej różnicy między pomiarami, n – liczba obserwacji. Wartości wskazane we wzorze (8) uzyskuje się podczas podstawowej analizy testem t Studenta dla danych zależnych.

Interpretując uzyskane wartości wielkości efektu, Cohen (1988) zaproponował następujące wartości: efekt słaby – (d) 0,20, efekt umiarkowany – (d) 0,50, efekt silny – (d) 0,80. Mimo że są one czytelne, w badaniach na przestrzeni lat pojawiała się rozbieżność w określaniu umiarkowanej wartości. Dlatego – jak zaznacza Sawilowsky (2009) – oprócz samego wyniku wskaźnika badacze powinni podawać przedziały, na podstawie których interpretują wielkości efektu. Lata pracy oraz większa świadomość statystyczna badaczy doprowadziły do opracowania rozszerzonej interpretacji wielkości efektu d . W analizach Monte-Carlo Sawilowsky (2009) wykazał bowiem, że wartości wielkości efektu d z łatwością mogą przekraczać 1,00, a czasami nawet 2,00, dla których brakuje interpretacji. W opracowaniu autor pokazał dokładniejszą propozycję interpretacji wielkości efektu: efekt bardzo słaby – (d) 0,10, efekt słaby – (d) 0,20, efekt umiarkowany – (d) 0,50, efekt silny – (d) 0,80, efekt bardzo silny – (d) 1,20, efekt ogromny – (d) 2,00 (Sawilowsky, 2009). Interpretacja ta pokrywa się z wcześniejszymi wskazaniem Cohena, ale została rozwinięta o nowe, które także pojawiają się w opracowaniach.

W badaniach skupionych na ocenie trzech i więcej grup czy pomiarów wykorzystanie współczynnika d Cohena może być utrudnione, szczególnie gdy warunków badawczych będzie znacznie więcej niż trzy. W takich planach badawczych stosuje się najczęściej analizę wariancji, gdzie w ocenie wielkości efektu można wykorzystać współczynnik eta-kwadrat (η^2), wyrażony wzorem (Olejnik, Algina, 2003, s. 435):

$$\eta^2 = \frac{SS_{\text{czynnik}}}{SS_{\text{cała}}} \quad (9),$$

gdzie: SS – suma kwadratów odchyłeń od średniej. Wartości wskazane we wzorze (9) są podstawowymi wyliczeniami prezentowanymi w analizie wariancji we wszystkich podstawowych pakietach statystycznych. Wielkość efektu η^2 jest dobrym współczynnikiem do porównań wewnątrz jednego planu badawczego w ocenie wielkości efektów głównych (Fritz i in., 2012). Zgłaszane są jednak zastrzeżenia do porównania wartości η^2 w analizie metadanych. Rozwiązaniem dla tego problemu może być wykorzystanie częściowego eta-kwadrat (η_p^2), który opisany jest wzorem (Olejnik, Algina, 2003, s. 435):

$$\eta_p^2 = \frac{SS_{\text{czynnik}}}{SS_{\text{czynnik}} + SS_{\text{bład}}} \quad (10),$$

gdzie: SS – suma kwadratów odchyłeń od średniej. Pokazane wartości składowe częściowej η^2 są także elementem podstawowego raportu analizy wariancji. Wzór ten może być wyrażony w postaci samych wartości uzyskanych w analizie wariancji (Baguley, 2009, s. 608):

$$\eta_p^2 = \frac{df_{\text{czynnik}} \times F_{\text{czynnik}}}{(df_{\text{czynnik}} \times F_{\text{czynnik}}) + df_{\text{bład}}} \quad (11),$$

gdzie: df – stopnie swobody, F – wartość analizy wariancji. Wskaźnik η^2 może być ograniczony przez zmiany błędu pomiaru, który wzrasta przy braku kontroli i identyfikacji zmiennych zakłócających. Wskazuje się, że nawet wykorzystanie częściowej η^2 między różnymi badaniami jest najbardziej uzasadnione, kiedy warunki i błędy pomiaru są porównywalne, jednak mimo to jest najczęściej wykorzystywanym wskaźnikiem wielkości efektu dla analizy wariancji (Fritz i in., 2012). Częściowa η^2 oraz podstawowa η^2 mogą być interpretowane podobnie: słaby efekt – (η^2) 0,01, umiarkowany efekt – (η^2) 0,06 oraz silny efekt – (η^2) 0,14 (Olejnik, Algina, 2000). Sapp (2006, za: Ferguson, 2009) sugeruje, że różnica między oszacowaniem η^2 i η_p^2 zmniejsza się wraz ze zwiększaniem wielkości próby.

W ocenie analizy wariancji w miejsce η^2 , która przeszacowuje proporcję wyjaśnianej wariancji, Hays i Winkler (1975, za: Olejnik, Algina, 2003) proponują wykorzystanie omega-kwadrat (ω^2), wyrażonej wzorem (Fritz i in., 2012, s. 11):

$$\omega^2 = \frac{\sigma_{\text{czynnik}}^2}{\sigma_{\text{czynnik}}^2 + \sigma_{\text{bład}}^2} \quad (12),$$

gdzie: σ – odchylenie standardowe dla populacji. Tutaj także – tak jak w przypadku podstawowego wzoru d Cohena – nie zawsze znamy wartości odchylenia w populacji. Musimy dokonać estymacji, sięgając po następujący wzór (Ferguson, 2009, s. 535):

$$\omega^2 = \frac{SS_{\text{czynnik}} - (k - 1)(S_{\text{bład}}^2)}{SS_{\text{cała}} + S_{\text{bład}}^2} \quad (13),$$

gdzie: SS – suma kwadratów odchyłeń od średniej, k – liczba grup, warunków badawczych dla czynnika, S^2 – wariancja. Znacznie łatwiejszy w opracowaniu jest wzór omega-kwadrat, który pozwala na wykorzystanie wartości analizy wariancji (Fritz i in., 2012, s. 12):

$$\omega^2 = \frac{(k - 1)(F_{\text{czynnik}} - 1)}{(k - 1)(F_{\text{czynnik}} - 1) + n} \quad (14),$$

gdzie: k – liczba grup, warunków badawczych dla czynnika, F – wartość analizy wariancji, n – całkowita liczba obserwacji. Prezentowane wzory eta i omega stanowią przykład dla prostego planu badawczego w obrębie efektu głównego. Próba oszacowania efektu interakcji w planach mieszanych czy analizach wieloczynnikowych wymaga wprowadzenia zmian, które można znaleźć w pracach Brzezińskiego (2021), Olejnik i Alginy (2003), Fritz i in. (2012) czy Fielda (2013). Wskaźnik ω^2 będzie wykazywał niższe wielkości efektu niż η^2 opracowany na tych samych danych (Olejnik, Algina, 2003), czego powodem jest przeszacowanie proporcji wyjaśnianej wariancji przez wskaźnik eta. Podobnie jednak jak dla różnic w wartościach wskaźników eta-kwadrat i częściowego eta-kwadrat, między omega-kwadrat i eta-kwadrat różnica oszacowania wielkości efektu będzie maleć zarówno przy zwiększaniu liczebności próby, jak i przy zmniejszaniu liczby czynników (Fritz i in., 2012). Wskaźnik omega-kwadrat jest rekomendowany w przypadku danych niezależnych (Brzeziński, 2021; Field, 2013). Interpretacja wartości omega-kwadrat pokrywa się z oceną wskaźnika eta: słaby efekt – (ω^2) 0,01, umiarkowany efekt – (ω^2) 0,06 oraz silny efekt – (ω^2) 0,14 (Olejnik, Algina, 2000).

2. Wielkość efektu testów nieparametrycznych

Oceny wielkości efektu powinny być prezentowane także w przypadku wykorzystania nieparametrycznych testów statystycznych. Pokazanie w takich przypadkach wartości d , g czy η^2 może nie być poprawne. Testy nieparametryczne wykorzystuje się z pewnych powodów, takich jak brak normalności rozkładu zmiennych czy jakościowa skala pomiaru (Cypryńska, Bedyńska, 2013). Porównanie znacznych asymetrycznych rozkładów za pomocą wartości średnich czy odchylenia standardowego może prowadzić do błędnej interpretacji rzeczywistej wielkości efektu. Podobnie będzie, kiedy zmienna zależna ma charakter porządkowy – wówczas średnia będzie nienajlepszą miarą tendencji centralnej. Wykorzystane nieparametryczne testy różnic mają określone wielkości efektu.

Test U Manna-Whitneya, dedykowany do porównania dwóch grup niezależnych, w ocenie wielkości efektu może zostać wsparty przez współczynnik korelacji dwuseryjnej Glassa (r_g), wyrażany wzorem (King, Minium, 2020, s. 566):

$$r_g = \frac{2(M_{rank1} - M_{rank2})}{n_1 + n_2} \quad (15),$$

gdzie: n – liczba obserwacji w grupie, M_{rank} – średnia rang w grupie. Wartości te można uzyskać podczas analizy testem Manna-Whitneya w podstawowych pakietach statystycznych, takich jak SPSS czy Statistica. Interpretacja wielkości efektu r_g jest zgodna z oceną przedziałów podstawowej korelacji r Pearsona, co oznacza: słaby efekt – (r) 0,00–0,30, umiarkowany efekt – (r) 0,31–0,50, silny efekt – (r) 0,51–0,70 oraz bardzo silny efekt – (r) 0,71–1,00.

Analiza różnic dwóch pomiarów zależnych testem Wilcozona w ocenie wielkości efektu wykorzystuje współczynnik korelacji dwuseryjnej dla par dopasowanych (r_c), opisany wzorem (King, Minium, 2020, s. 572):

$$r_c = \frac{4|T - (\frac{R_+ + R_-}{2})|}{n(n + 1)} \quad (16),$$

gdzie: n – liczba wszystkich obserwacji, T – suma rang o wyższej wartości (R_+ lub R_-), R_+ – suma rang dodatnich, R_- – suma rang ujemnych. Wykorzystanie wzoru (16) może być utrudnione koniecznością zestawienia par wyników i samodzielnego wyliczenia rang. Korzystając z pakietów statystycznych, znacznie szybciej można wskazać wielkość efektu r_c za pomocą uproszczonego wzoru (Rosnow, Rosenthal, 2003, s. 231):

$$r_c = \frac{z}{\sqrt{n}} \quad (17),$$

gdzie: z – wartość statystyki testowej, n – liczba wszystkich obserwacji. Uzyskując te dane z analizy statystycznej, łatwiej wskazać wartość wielkości efektu r_c , której interpretacja jest zgodna z interpretacją współczynnika r Pearsona.

Badanie porównujące trzy i więcej grup niezależnych testem nieparametrycznym stwarza najczęściej podstawę do wykorzystania testu H Kruskala-Walisa. Wielkością efektu dla testu H jest statystyka epsilon-kwadrat (ε^2), opisywana wzorem (King, Minium, 2020, s. 572):

$$\varepsilon^2 = \frac{\chi_H^2}{(n^2 - 1)/(n^2 + 1)} \quad (18),$$

gdzie: n – liczba wszystkich obserwacji, χ_H^2 – wartość statystyki testowej. Wartość testu Kruskala-Wallisa także jest możliwa do uzyskania przy wykorzystaniu pakietów statystycznych. Wyniki wielkości efektu ε^2 można interpretować jako: słaby efekt – (ε^2) 0,01, umiarkowany efekt – (ε^2) 0,06 oraz silny efekt – (ε^2) 0,14 (Olejnik, Algina, 2000).

Test Friedmana stosowany jest do oceny trzech i/lub więcej pomiarów zależnych jako nieparametryczny odpowiednik analizy wariancji z powtarzaniem pomiarem. W analizie tej do oceny wielkości efektu można wykorzystać współczynnik Kendalla (W), wyrażony wzorem (King, Minium, 2020, s. 578):

$$W = \frac{\chi_F^2}{n(k-1)} \quad (19),$$

gdzie: χ_F^2 – wartość statystyki testowej, n – liczba wszystkich obserwacji, k – liczba porównanych pomiarów. Wartość W Kendalla jest automatycznie wskazywana przez większość pakietów statystycznych, a jego rozpiętość mieści się w przedziale od 0,00 (brak efektu) do 1,00 (całkowity efekt).

3. Wielkość efektu dla zmiennych jakościowych

Wiele badań w swoim planie zakłada nominalny pomiar. Dla zestawu dwóch zmiennych, które z natury są dychotomiczne lub zostały poddane dychotomizacji, w ocenie wielkości efektu zamiast istotności samego testu *chi*-kwadrat (χ^2) można wykorzystać opisane niżej wskaźniki. W analizie dwóch zmiennych o dwóch wartościach (w planie 2x2) zastosowanie znajdzie wskaźnik ϕ (φ), opisany wzorem (Lourel, Gueguen, Pascaul, Mouda, 2011, s. 632):

$$\varphi = \sqrt{\frac{\chi^2}{n}} \quad (20),$$

gdzie: χ^2 – wartość statystyki testowej *chi*-kwadrat, n – liczba wszystkich obserwacji. Wskaźnik φ może przyjmować wartości od 0,00 (brak związku) do 1,00 (związek doskonały). King i Minium (2020) zaznaczają, że jeżeli w tabeli porządkujemy niższą wartość zmiennej jako zero, a wyższą jako jeden, to wartość wskaźnika φ będzie równa wartości korelacji r Pearsona. Wartość ϕ należy interpretować jako: słaby efekt – (φ) 0,10, umiarkowany efekt – (φ) 0,30 oraz silny efekt – (φ) 0,50 (Brzeziński, 2021; King, Minium, 2020).

Analiza dwóch dwukategorialnych zmiennych oprócz wskaźnika φ może opierać się na ocenie ilorazu szans, ryzyku względnym czy ryzyku różnicy. W tabeli 1 zamieszczono przykładowy schemat planu badawczego wykorzystanego

do omówienia kolejnych wzorów. Analiza wystąpienia zdarzeń w prostym planie eksperymentalnym daje możliwość wykorzystania ilorazu szans (OR), który można wyrazić wzorem (Ferguson, 2009, s. 535):

$$OR = \frac{A/B}{C/D} \quad (21),$$

gdzie duże litery oznaczają liczbę obserwacji w planie badawczym opisanym szczegółowo w tabeli 1. Iloraz szans należy interpretować jako szansę zajścia zdarzenia w jednej grupie w stosunku do szansy jej zajścia w drugiej grupie. Analizując wyniki OR , można się spotkać z trzema możliwymi wynikami:

- a) $OR = 1$ – brak wpływu warunków badania na wystąpienie zdarzenia. W obu grupach: kontrolnej i eksperymentalnej – szansa na zdarzenie jest podobna bądź taka sama, co oznacza, że manipulacja eksperymentalna nie wpłynęła na występowanie zdarzenia,
- b) $OR > 1$ – widać wpływ oddziaływania eksperymentalnego. Im wyższy niż 1,00 wynik OR , tym większa szansa na wystąpienie w grupie eksperymentalnej zdarzenia. Na przykład wartość $OR = 2,00$ oznacza, że szansa na wystąpienie zdarzenia jest dwa razy większa w grupie eksperymentalnej,
- c) $OR < 1$ – wynik sugerujący większą szansę na zdarzenie w grupie kontrolnej. Osiągnięcie takiego wyniku może dziwić, ale jest on zupełnie normalny w sytuacji braku wyróżnienia grupy podstawowej w badaniu o charakterze eksploracyjnym (Brzeziński, 2021).

Tabela 1. Przykład podstawowego planu badawczego dwóch dychotomicznych zmiennych

Grupa	Badane zdarzenie	
	wystąpienie zdarzenia	brak zdarzenia
Eksperymentalna	A	B
Kontrolna	C	D

Uwaga: duże litery reprezentują liczbę obserwacji w każdym warunku badania.

Źródło: opracowanie własne na podstawie (Brzeziński, 2021, s. 226).

Rosnow i Rosenthal (2003) zaznaczają, że wartość OR ma tendencję do osiągnięcia nieskończoności w sytuacji badawczej, kiedy mamy do czynienia z bardzo małą liczbą wartości mianownika (małym wynikiem dla grupy kontrolnej, podstawowej), co może utrudniać stawianie praktycznych wniosków. Propozycją dodatkowej interpretacji ilorazu szans jest: słaby efekt – (OR) 1,50, umiarkowany efekt – (OR) 2,00 oraz silny efekt – (OR) 3,00 (Sullivan, Feinn, 2012). Inną możliwością oceny planów badawczych 2x2 jest ryzyko względne (RR), opisane wzorem (Ferguson, 2009, s. 535):

$$RR = \frac{\left(\frac{A}{A+B}\right)}{\left(\frac{C}{C+D}\right)} \quad (22),$$

gdzie duże litery oznaczają liczbę obserwacji w planie badawczym opisanym szczegółowo w tabeli 1. Wskaźnik ryzyka względnego bardzo często stosowany jest w ocenie danych biomedycznych w odniesieniu do analiz czynników ryzyka wystąpienia zjawisk chorobowych. Wynik RR mówi o wzroście prawdopodobieństwa wystąpienia zdarzenia w grupie eksperymentalnej w stosunku do grupy kontrolnej. Wynik poniżej 1,00 należy interpretować jako mniejsze ryzyko na wystąpienie zdarzenia w grupie eksperymentalnej. Im wyższa od 1,00 wartość RR , tym większe prawdopodobieństwo wystąpienia zdarzenia w grupie eksperymentalnej w stosunku do grupy kontrolnej. Tak jak w przypadku OR , wynik RR równy 1,00 sugeruje podobne lub takie samo prawdopodobieństwo wystąpienia zdarzeń w grupach. Inną możliwą interpretacją ryzyka względnego jest: słaby efekt – (RR) 2,00, umiarkowany efekt – (RR) 3,00 oraz silny efekt – (RR) 4,00 (Sullivan, Feinn, 2012). Kolejną możliwością oceny efektu dla planów badawczych 2x2 jest ryzyko różnic (RD), opisane wzorem (Ferguson, 2009, s. 535):

$$RD = \frac{A}{A+B} - \frac{C}{C+D} \quad (23),$$

gdzie duże litery oznaczają liczbę obserwacji w planie badawczym opisanym szczegółowo w tabeli 1. Ryzyko różnicy jest znacznie prostsze w interpretacji. Jego wynik świadczy o faktycznej różnicy między grupami (Ferguson, 2009). Dla przykładu RD równe 0,06 oznacza 6% ryzyka różnicy między grupami kontrolną i eksperymentalną dla wystąpienia zdarzenia. Jak zaznacza Ferguson (2009), praktyczna interpretacja ryzyka różnicy może być jednak w warunkach badań klinicznych trudniejsza; RD może okazać się nieistotne w sytuacji ryzyka 75% w grupie kontrolnej i 81% w grupie eksperymentalnej, ale w odniesieniu do porównania ryzyka 1% w grupie kontrolnej i 7% w grupie eksperymentalnej może wykazywać większą istotność (Ferguson, 2009).

Dla większych planów badawczych dwóch zmiennych, zakładających istnienie więcej niż dwóch kategorii (plan 2x3 i większe), zastosowanie znajdzie przekształcony wskaźnik Cramera (V), opisany wzorem (Brzeziński, 2021, s. 226):

$$V = \sqrt{\frac{\chi^2}{n(df_{mniejsze})}} \quad (24),$$

gdzie: χ^2 – wartość statystyki testowej *chi*-kwadrat, n – liczba wszystkich obserwacji, $df_{mniejsze}$ – mniejsza liczba stopni swobody dla kolumn lub dla wierszy. Jak zaznaczają King i Minium (2020), wskaźnik V można wykorzystać dla każdej tabeli kontyngencji. W miarę zwiększania kategorii zmiennych wynik V może być coraz trudniejszy w praktycznej interpretacji analizowanych danych. Wartość wyniku V należy interpretować podobnie jak podstawowy wskaźnik φ : słaby efekt – (V) 0,10, umiarkowany efekt – (V) 0,30 oraz silny efekt – (V) 0,50 (Brzeziński, 2021; King, Minium, 2020).

Wskazane w tej części wzory wskaźników wielkości efektu, ich dopasowanie do wykorzystanych testów statystycznych oraz propozycje ich interpretacji zestawiono w tabeli 2. Wybór odpowiedniego wskaźnika wielkości efektu powinien być traktowany równie poważnie jak określenie zgodnego z metodologią badania kierunku statystycznej analizy danych (Pek, Flora, 2018).

Tabela 2. Dopasowanie wskaźników wielkości efektu oraz przedziały interpretacji

Test	Wskaźnik wielkości efektu	Interpretacja wielkości efektu
<i>T</i> Studenta	d Cohena g Hedgesa Δ Glassa	0,20 – słaby 0,50 – umiarkowany 0,80 – silny ^a 0,10 – bardzo mały 0,20 – mały 0,50 – umiarkowany 0,80 – silny 1,20 – bardzo silny 2,00 – ogromny ^b
<i>U</i> Manna-Whitneya	r_g	0,00–0,30 – słaby 0,31–0,50 – umiarkowany 0,51–0,70 – silny 0,71–1,00 – bardzo silny
Wilcoxona	r_c	
Test Friedmana	W Kendalla	
Analiza wariancji	$\eta^2, \eta_p^2, \omega^2$	0,01 – słaby 0,06 – umiarkowany 0,14 – silny
<i>H</i> Kruskala-Wallisa	ε^2	
<i>Chi</i> -kwadrat	φ, V	0,10 – słaby 0,30 – umiarkowany 0,50 – silny

Objaśnienie: ^a – propozycja interpretacji Cohena; ^b – rozszerzona interpretacja Sawilowsky'ego.

Źródło: opracowanie własne.

ZAKOŃCZENIE

Dobór odpowiedniej ścieżki wnioskowania statystycznego oraz przeprowadzenie najbardziej zaawansowanych analiz powinny stwarzać szansę na budowanie wniosków wykraczających poza binarne systemy odpowiedzi na pytania badawcze. W części empirycznej opracowania zaproponowano, aby statystyczna analiza danych obok wartości poziomu istotności stawiała wartość wielkości efektu wraz z przedziałami ufności. Odpowiednio dobrane do modelu badawczego wskaźniki wielkości efektu powinny zostać poddane interpretacji. Co więcej, wartości wskazane w analizie danych, w tym wielkości efektu, powinny być opatrzone komentarzem dotyczącym praktycznych implikacji. Do porównania dwóch grup czy pomiarów można zastosować w analizie testami parametrycznymi wskaźnik d Cohena, g Hedgesa lub Δ Glassa, a w analizie testami nieparametrycznymi – współczynnik korelacji rangowej r_s Glassa lub współczynnik korelacji dwuseryjnej par dopasowanych r_c . W ocenie trzech i więcej grup czy pomiarów w analizie wariancji można wykorzystać etę-kwadrat (η^2) lub omegę-kwadrat (ω^2). Nieparametryczne porównanie trzech i więcej grup/pomiarów daje szansę na ocenę wielkości efektu wskaźnikami epsilon-kwadrat (ϵ^2) oraz W Kendalla. Analiza wielkości efektu danych jakościowych wykorzystuje współczynnik f_i (ϕ), V Cramera, iloraz szans (OR) oraz ryzyko względne (RR). Wartości poziomu istotności nadal pozostają ważnym elementem analizy danych, ale trzeba pamiętać, że „Bóg tak samo kocha 0,060, jak 0,050” (Rosnow, Rosenthal, 2003, s. 1277, za: Henson, 2006, s. 605).

BIBLIOGRAFIA

LITERATURA

- American Psychology Association (APA). (1994). *Publication Manual of the American Psychology Association* (4th ed.). Washington: Author.
- American Psychology Association (APA). (2020). *Publication Manual of the American Psychology Association* (7th ed.). Washington: Author.
- Baguley, T. (2009). Standardized or Simple Effect Size: What Should Be Reported? *British Journal of Psychology*, 100(3), 603–617. DOI: 10.1348/000712608X377117
- Bosco, F.A., Aguinis, H., Singh, K., Pierce, C.A. (2015). Correlational Effect Size Benchmarks. *Journal of Applied Psychology*, 100(2), 431–439. DOI: 10.1037/a0038047
- Brzeziński, J.M. (2021). Testy istotności różnic i wskaźniki wielkości efektu ES – wybrane zagadnienia. W: *Metodologia badań psychologicznych* (s. 205–234). Warszawa: Wydawnictwo Naukowe PWN.
- Cohen, J. (1988). *Statistical Power: Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates.
- Cypryańska, M., Bedyńska, S. (2013). Testy t-Studenta i ich nieparametryczne odpowiedniki. W: S. Bedyńska, M. Cypryańska (red.), *Statystyczny drogowskaz 1. Praktyczne wprowadzenie do wnioskowania statystycznego* (s. 159–193). Warszawa: Wydawnictwo Akademickie Sedno.

- Ferguson, C.J. (2009). An Effect Size: A Guide for Clinicians and Researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538. DOI: **10.1037/a0015808**
- Field, A. (2013). *Discovering Statistics Using IBM SPSS*. Thousand Oaks: Sage.
- Fritz, C.O., Morris, P.E., Richler, J.J. (2012). Effect Size Estimates: Current Use, Calculations and Interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. DOI: **10.1037/a0024338**
- Grissom, R.J., Kim, J.J. (2011). *Effect Sizes for Research: Univariate and Multivariate Applications*. New York: Routledge, Taylor and Francis Group. DOI: **10.4324/9780203803233**
- Henson, K. (2016). Effect-Size Measures and Meta-Analytic Thinking in Counseling Psychology Research. *Counseling Psychology*, 34(5), 601–629. DOI: **10.1177/00110005283558**
- King, B.M., Miniun, E.W. (2020). *Statystyka dla psychologów i pedagogów*. Warszawa: Wydawnictwo Naukowe PWN.
- Kirk, R.E. (2001). Promoting Good Statistical Practices: Some Suggestions. *Educational and Psychological Measurement*, 61(2), 213–218. DOI: **10.1177/00131640121971185**
- Kühberger, A., Fritz, A., Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. *PLoS ONE*, 9(9), 1–8. DOI: **10.1371/journal.pone.0105825**
- Lourel, M., Gueguen, N., Pascaul, A., Mouda, F. (2011). The Effect-Size: A Simple Methodology for Determining and Evaluating the “Effect-Size”. *Psychology*, 2(6), 631–632. DOI: **10.4236/psych.2011.26096**
- Olejnik, S., Algina, J. (2000). Measures of Effect Size for Comparative Studies: Applications, Interpretations, and Limitations. *Contemporary Educational Psychology*, 25(3), 241–286. DOI: **10.1006/ceps.2000.1040**
- Olejnik, S., Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, 8(4), 434–447. DOI: **10.1037/1082-989X.8.4.434**
- Pek, J., Flora, D.B. (2018). Reporting Effect Size in Original Psychological Research: A Discussion and Tutorial. *Psychological Methods*, 23(2), 208–225. DOI: **10.1037/met0000126**
- Rosenthal, R., Rosnow, R.L., Rubin, D.B. (2000). *Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach*. Cambridge: Cambridge University Press. DOI: **10.1017/CBO9780511804403**
- Rosnow, R.L., Rosenthal, R. (2003). Effect Size for Experimenting Psychologists. *Canadian Journal of Experimental Psychology*, 57(3), 221–237. DOI: **10.1037/h0087427**
- Sawilowsky, S.S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599. DOI: **10.22237/jmasm/1257035100**
- Sullivan, G.M., Feinn, R. (2012). Using Effect Size – or Why the *p* Value Is Not Enough. *Journal of Graduate Medical Education*, 4(3), 279–282. DOI: **10.4300/JGME-D-12-00156.1**
- Sun, S., Pan, W., Wang, L.L. (2010). A Comprehensive Review of Effect Size Reporting and Interpreting Practice in Academic Journals in Education and Psychology. *Journal of Educational Psychology*, 102(4), 989–1004. DOI: **10.1037/a0019507**
- Vacha-Haase, T., Thompson, B. (2004). How to Estimate and Interpret Various Effect Sizes. *Journal of Counseling Psychology*, 51(4), 473–481. DOI: **10.1037/0022-0167.51.4.473**

NETOGRAFIA

- Coe, R. (12–14.09.2002). *It's the Effect Size, Stupid*. British Educational Research Association Annual Conference, Exeter, Great Britain. Pobrane z: <https://f.hubspotusercontent30.net/hubfs/5191137/attachments/ebe/ESguide.pdf>

ABSTRACT

Reflecting the changing statistical practice in psychological research, dominated by null hypothesis testing using a decision about the level of significance of the results, the recommendations are indicated for reporting effect sizes in papers. The study presents the concept of the effect size and indicates its place in data analysis regarding to outcome's significance. The purpose of the work is to describe selected effect size indicators and to point the need of use and their proper presentation and interpretation in social sciences empirical work data analysis reports. Considering statistical analysis approach limits based on significance level only, the study presents the possibility of including in the data analysis an indicator of a more practical use which is the size of the effect. By using the most popular analysis methods, such as, Student t -test, univariate analyses of variance in between- and within-group schemes as well as Wilcoxon test, Mann-Whitney's U , Kruskal-Wallis H , Friedman's test and considering analysis for qualitative data, matched to research plans indicators of the effect size were presented. The paper presents the use, calculation and interpretation of the size effect such as: Cohen's d , Hedges g , delta, Glass's r_g , matched pairs correlation r_c , eta-square, omega-square and epsilon-square, Kendall's W and f_t , Cramer's V as well as odds ratio and relative risk. The presentation of the effect size indicators was contrasted with the corresponding research plans and the type of data collected.

Keywords: effect size; null hypothesis; statistical analysis; psychological research

