

JERZY MARIAN BRZEZIŃSKI

CO TO ZNACZY,
ŻE WYNIKI PRZEPROWADZONYCH PRZEZ PSYCHOLOGÓW
BADAŃ NAUKOWYCH
PODDAWANE SĄ ANALIZIE STATYSTYCZNEJ?

Autor artykułu, odwołując się do współczesnego stanu świadomości (*awareness*) metodologicznej psychologów oraz własnego wieloletniego doświadczenia badawczego i edukacyjnego, i inspirowany takimi pracami, jak: Cohen (1990-2006, 1994/2006) i Wilkinson & Task Force on Statistical Inference (1999), podejmuje węzłowe problemy stosowania metod analizy statystycznej danych pochodzących z badań psychologicznych. Zwraca też uwagę na możliwe źródła nadużyć oraz optymalne rozwiązania, których przyjęcie przez badaczy będzie minimalizowało wariancję błędu (*error variance*) i przyczyni się do znaczącego obniżenia kreowanych artefaktów. Celem artykułu jest też zwrócenie uwagi na możliwe źródła naukowych nadużyć (*scientific misconduct*). Autor skupił się na następujących zagadnieniach: standardowym ujęciu procesu badawczego, pojęciu istotności, miejscu analizy statystycznej danych (*data*) w procesie badawczym, konieczności replikacji wyników badań naukowych, opozycji: NHST vs *confidence interval*, ważności ustalania w badaniach psychologicznych *effect size* oraz celowości prowadzenia *multivariate analysis*.

Słowa kluczowe: proces badawczy, świadomość metodologiczna, wielkość efektu (ES), NHST, przedziały ufności, replikacja, wielozmiennowa analiza.

„Może by zaorać co...”

Sławomir Mrożek, *Indyk*

I. UWAGI WSTĘPNE

Zacznę od wyrażenia mojego przekonania, że badanie naukowe (a w przypadku psychologii ma ono charakter **empiryczny**) nie polega na wrzucaniu do komputera różnych wyników i dopasowywaniu do nich testów istotności oraz miar korelacji tak długo, aż „coś” z tego wyjdzie. Jeżeli uzyskamy, przynajmniej z części wyników, jakieś istotne różnice czy związki korelacyjne, to wówczas skupimy się właśnie na nich i spróbujemy je „zrozumieć”. Kiedyś, w epoce przedkomputerowej (kto jeszcze pamięta te czasy?), taka strategia badawcza była mało efektywna, ale dziś, w dobie „panowania” pakietów statystycznych, możemy bardzo szybko porównać wszystko z wszystkim i zobaczyć, co wyszło. Coraz bardziej doskonałe (szybkie!) komputery i coraz wnikliwsze (ale też prostsze w obsłudze, przyjazne) statystyczne programy komputerowe sprawiają, że dziś student już na pierwszym roku studiów sięga po procedury, które przed laty były studentom niedostępne. Ktoś powie: postęp. Tak, ale czy to nie jest jakoś tak, jak z owym psem, którym kręci jego ogon? Tym, że nie zawsze użytkownik pakietu ma świadomość poprawności (w sensie zgodności materiału empirycznego z założeniami zastosowanej metody) użycia jakiejś metody statystycznej, raczej nikt sobie specjalnie głowy nie zawraca. Ważne, żeby „coś” wyszło i żeby można to było próbować opublikować. Takie postępowanie przypomina pisanie powieści od ostatniego rozdziału.

Oczywiście przerysowałem problem, ale to wcale nie oznacza, że problemu nie ma. Problemem jest bowiem powstawanie coraz większej liczby prac przyziemnie empirycznych, bez oddechu teoretycznego, bez wizji. Są to prace o niczym, śmieciowe, ale – bywa – poprawnie „porachowane”. Mam poczucie, że cofamy się do okresu wczesnego behawioryzmu i prymitywnego operacjonizmu, który negował wartość poznawczą teorii (to B. F. Skinner, 1950, napisał słynne zdanie: *Theories are fun*). W efekcie tacy badacze kierują się swoistą strategią (King, Minium, 2009, s. 357): „śmieci włóż i śmieci weź”.

Zatem pisząc o analizie statystycznej wyników badań psychologicznych, nie mogę nie zacząć „od początku”. Badanie empiryczne bowiem, a w nim zawartą analizę statystyczną wyników, musi poprzedzać refleksja teoretyczna. Badanie służy falsyfikacji (czy konfirmacji) hipotez, które traktują o zmiennych teoretycznych, które badacz w taki albo inny sposób zoperacjonalizował. Te zaś zmienne zostały uprzednio zdefiniowane na gruncie jakiejś teorii (ta zaś ma cha-

rakter empiryczny) i jakiegoś paradygmatu (mówiąc górnolotnie językiem Thomasa Kuhna). I w tym sensie badanie, w którym analizowane są – powiedzmy – tylko dwie zmienne, nie jest gorsze od badania analizującego dwadzieścia zmiennych, o ile to pierwsze skupia uwagę na czymś, co jest rzeczywiście ważne. Doniosłość teoretyczna badania nie liczbą zmiennych jest mierzona. Dobrym przykładem jest badanie przeprowadzone przed ponad 100 laty przez Yerkesa i Dodsona, które – mimo „skromności” – do dziś jest cytowane w wielu podręcznikach psychologii na całym świecie.

Od wielu już lat zajmuję się dydaktyką z zakresu metodologii i zastosowań statystyki w badaniach społecznych – na poziomie magisterskim i doktoranckim. Mam na ten temat swoje obserwacje i przemyślenia. Mam też swoje upodobania (np. do ANOVA). Śledzę literaturę przedmiotu oraz staram się – z myślą o studentach i doktorantach – upowszechniać kluczowe prace z zakresu metodologii, psychometrii i zastosowań statystyki. Z mojej inicjatywy przetłumaczono na język polski dwa, jak mi nie mam – dobre, podręczniki statystyki adresowane do psychologów i pedagogów (por. Ferguson, Takane, 2003; King, Minium, 2009). Ten ostatni – co warto zauważyć – uwzględnia już wytyczne specjalnego zespołu ekspertów powołanego przez APA (por. Wilkinson i Task Force on Statistical Inference, 1999), który opracował zalecenia dla psychologów przygotowujących prace z wykorzystaniem analizy statystycznej danych (m.in. nacisk położono na: wskaźniki wielkości efektu, przedziały ufności jako alternatywę dla testowania hipotez zerowych oraz analizę mocy testu statystycznego¹ (także nowszy raport: JARS, 2008). Wydałem też zbiory tekstów z tego zakresu.

Szczęśliwie się składa, że w ostatnich latach ukazały się także i polskie, dobre opracowania dotyczące AS, pisane przez psychologów (a więc autorzy rozumieją potrzeby odbiorców). Mam na myśli zwłaszcza trzy opracowania: Grażyny Wieczorkowskiej i Jerzego Wierzbińskiego (2011) pt. *Statystyka. Analiza badań społecznych*; Piotra Francuza i Roberta Mackiewicza (2005) pt. *Liczy nie wiedzą, skąd pochodzą. Przewodnik po metodologii i statystyce nie tylko dla psychologów*; Sylwii Bedyńskiej i Anety Brzezickiej (2007) pt. *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii* (z bogatą ilustracją zastosowań pakietu SPSS, wersja 14, do analizy danych psychologicznych).

¹ Dla przypomnienia: **moc testu** (*power of the test*) jest prawdopodobieństwem odrzucenia hipotezy zerowej, gdy *de facto* jest ona fałszywa. Ponieważ prawdopodobieństwo nieodrzućenia fałszywej hipotezy zerowej wynosi β , zatem prawdopodobieństwo podjęcia przez badacza poprawnej decyzji odnośnie do fałszywej H_0 wynosi: $1 - \beta$.

Ulubionym przez psychologów pakietem statystycznym jest SPSS. Warto tedy wydać tłumaczenie kapitalnego opracowania pióra Andy'ego Fielda (2009): *Discovering statistics using SPSS* (wyd. 3) – oparty na wersji 16 i 17 SPSS – czy: Christine P. Dancey i Johna Reidy (2011): *Statistics without Maths for psychology* (wyd. 5) – oparta na wersji 18 IBM®SPSS®Statistics. Brakuje mi też opracowań, które wprowadziłyby do środowiska psychologicznego drugi pakiet statystyczny: STATA (por. www.stata.com).

Środowisku psychologicznemu, zwłaszcza doktorantom, potrzebny jest dobry podręcznik zaawansowanej analizy statystycznej (wielozmiennowej). Mój wybór pada na opracowanie Barbary G. Tabachnick oraz Lindy S. Fidel (2001): *Using multivariate statistics* (966 stron dużego formatu!). Pojawia się jednak zasadnicze pytanie: kto go przełoży i kto go wyda?

Spróbuję więc podzielić się ze swoimi wątpliwościami i preferencjami. Liczę na to, że specjaliści o dużym doświadczeniu badawczym, którzy zechcieli przyjąć Redakcji i moje zaproszenie do udziału w tej dyskusji, zwrócą uwagę (nastąpi kumulacja wielkości efektu!) psychologów (zwłaszcza tych, którzy dopiero wkraczają na fascynującą drogę odkrywania nieznanego i doskonalenia narzędzi poznania) na właściwe stosowanie metod statystycznej analizy danych. Rzecz jasna mój szkic nie podejmuje wszystkich wątków; do tego potrzebna byłaby nie tak mała monografia. Starałem się jednak odsyłać do prac, które kiedyś czytałem (czytam) i do których często wracam. To zaplecze literaturowe może okazać się pomocne dla Koleżanek/Kolegów zaczynających dialog z psychologią naukową, a z metodologią i statystyką w szczególności.

Analiza statystyczna, szeroko pojmowana, obejmuje nie tylko „oglądanie” danych empirycznych w sposób usystematyzowany i ich opis czy testowanie hipotez w modelu NHST (od ang. *Null Hypothesis Statistical Testing*), lecz także analizę mocy testów, problematykę projektowania sposobu wyłaniania próby z danej populacji, konstruowania modelu statystycznego badania (np. ANOVA/MANOVA w odniesieniu do eksperymentu czy MCR (od ang. *Multiple Correlation/Regression*) w odniesieniu do badań korelacyjnych), procedur statystycznej kontroli zmiennych ubocznych i zakłócających, konstruowania przedziałów ufności, statystyczną stronę operacjonalizacji zmiennych i interpretacji wyników testów psychologicznych, ocenę istotności zmiennych niezależnych za pomocą wskaźników wielkości efektu (*effect size*), trafności zewnętrznej (zakres możliwych uogólnień uzyskanych rezultatów badawczych) itd.

Aby sensownie o niej rozmawiać, musimy wpierw ją odnaleźć w strukturze procesu badawczego. I od tego zacznę mój artykuł.

II. STANDARDOWE UJĘCIE PROCESU BADAWCZEGO W PSYCHOLOGII

1. *Cechy poznania naukowego*

Jedną z konstytutywnych cech poznania naukowego (obok **intersubiektywności**² i **powtarzalności**³), odróżniającą je od innych rodzajów poznania, jest ta, która wskazuje na jego **idealizacyjny charakter**⁴. Jeżeli zatem odwołamy się do bliskiego mi ogólnego modelu poznania naukowego, jaki oferuje **Idealizacyjna Teoria Nauki (ITN)**⁵, to postępowanie typowego badacza wygląda – w wielkim skrócie – tak, że dokonuje on swoistej, myślowej deformacji rzeczywistości (i nie ma znaczenia, jaka to jest rzeczywistość), budując jej uproszczony, a więc w jakimś sensie karykaturalny obraz – tworząc jej **idealizację**. To, przy przyjęciu określonych **założeń idealizujących**, można – właśnie przez oderwanie się od szczegółów – formułować **twierdzenia idealizacyjne** traktujące o zachodzeniu jakiejś zależności opisanej **prawem nauki**. Rzecz jasna twierdzenia te nie odnoszą się do obiektów materialnych występujących w warunkach rzeczywistości zachodzących w otaczającym nas świecie. Wiemy bowiem dobrze, że na przykład wahadła matematyczne nie istnieją realnie. Chcąc tedy przybliżyć nasze twierdzenie idealizacyjne do rzeczywistości, musimy poddać je swoistej procedurze – odwrotnej do procedury idealizacji, a mianowicie musimy je **skonkretyzować**. Dokonuje się to poprzez uchylanie kolejnych założeń idealizujących w procesie stopniowej konkretyzacji. Postępując konsekwentnie, otrzymamy serię modeli coraz bardziej adekwatnie przystających do rejestrowanych danych. Sekwencja stworzonych w ten sposób modeli (przez kolejne konkretyzacje wyjściowego, najbardziej abstrakcyjnego, modelu podstawowego) tworzy **teorię idealizacyjną**.

2. *Operacjonalizacja terminu: „istotność”*

Jednym z podstawowych pojęć ITN jest pojęcie **istotności** (por. Buczkowski, Nowak, 1979). W trakcie rozwoju ITN, a zwłaszcza gdy próbowano konfronto-

² W sensie: Ajdukiewicz (1983).

³ W sensie: Wojciszke (2006, 2011), Rosenthal (1991), Neuliep (1991).

⁴ W sensie: Nowak, 1980; Nowakowa, Nowak, 2000; Brzeziński, 1982; Gaul, 1990; Krajewski, 1998.

⁵ W sensie: Nowakowa, Nowak, 2000; w odniesieniu do nauk behawioralnych: Brzeziński, 1982; Gaul, 1990.

wać ITN z praktyką badawczą różnych szczegółowych nauk, wypracowano jej bardziej specyficzne **parafrazy** (por. Egiert, 2000; także w odniesieniu do nauk behawioralnych, a więc i psychologii, por. koncepcję **protoidealizacji**: Brzeziński, 1982; Gaul, 1990). To na jej gruncie nadaje się pojęciu istotności specyficzny sens empiryczny, wiążący je z proporcją zmienności zmiennej zależnej tłumaczoną (wyjaśnianą) przez określoną zmienną niezależną (zmienne niezależne i ich możliwe interakcje).

Zatem jednym z najważniejszych pytań, na które musi udzielić odpowiedzi badacz przeprowadzający badanie empiryczne, jest pytanie o istotność jakiejś zmiennej niezależnej (lub pewnej liczby zmiennych i ich interakcji) dla określonej zmiennej zależnej. Pochodną tego pytania – gdy badacz rozpatruje wpływ większej (co najmniej 2) liczby zmiennych – jest to, które dotyczy porządku istotnościowego zmiennych zakwalifikowanych przez badacza jako istotne dla danej zmiennej zależnej.

Ponieważ rozpatrywane w psychologii zależności – wiążące daną zmienną (to jej teorię, „mniejszą” czy „większą”, będzie konstruował badacz) ze zmiennymi uznanymi przez badacza za istotne dla niej – mają charakter **statystyczny**, więc staje się oczywiste, że badacze wiążą pojęcie istotności jakiejś zmiennej X (potraktowanej jako zmienna niezależna, przyczyna) dla jakiejś innej zmiennej, powiedzmy Y (potraktowanej jako zmienna zależna, skutek), z obserwowaną **zmiennością** Y, będącą efektem oddziaływania na nią właśnie tej wyróżnionej przez badacza zmiennej X_j , która – wedle jego wiedzy – może być przyczyną zmiennej Y (będącej skutkiem). Znaną w statystyce i chętnie przez badaczy stosowaną miarą zmienności jest **wariancja** – zwłaszcza z uwagi na jej własności.

Wszystkie zmienne niezależne, uznane przez badacza za wywierające wpływ na zmienną Y, a więc – na gruncie jego aktualnej wiedzy – **istotne** dla niej, można podzielić na dwie klasy, przyjmując jako kryterium **kontrolowanie** przez badacza, w danym badaniu, ich wpływu na zmienną Y. Pierwsza klasa obejmuje **zmienne kontrolowane** – są one źródłem **wariancji kontrolowanej** zmiennej Y (inaczej: **wyjaśniają** określony procent jej wariancji całkowitej). Druga zaś klasa obejmuje **zmienne niekontrolowane** – są one źródłem **wariancji niekontrolowanej** zmiennej Y (inaczej zwanej wariancją resztową albo wariancją błędu).

Chcąc więc poznać wielkość wpływu zmiennej X na zmienną Y, trzeba – w kontrolowanych warunkach badania empirycznego – określić procentowy udział wariancji cząstkowej, której źródłem jest dana zmienna, w wariancji całkowitej zmiennej Y.

Powiązanie pojęcia istotności zmiennej X z miarą zmienności wyników zmiennej zależnej Y, którą jest wariancja zmiennej Y, nadaje pojęciu istotności

sens statystyczny. Psycholog, na przykład mówiąc, że „IQ” (zmienna niezależna) jest zmienną istotną dla zmiennej „osiągnięcia szkolne” (zmienna zależna), ma na myśli to, że zmienna „IQ” **wyjaśnia** określony, wyższy od pewnej wartości progowej, procent wariancji całkowitej zmiennej „osiągnięcia szkolne” (ten procent można precyzyjnie ustalić) i że jest ona jednym ze źródeł (gdyż – na ogół – zbiór zmiennych istotnych dla Y nie jest zbiorem jednostkowym) wariancji wyjaśnionej zmiennej Y. Zmienne, których wpływy na Y, w danym badaniu, nie są przez badacza kontrolowane, są źródłem wariancji **niekontrolowanej, niewyjaśnionej.**

3. Miejsce AS w strukturze procesu badawczego w psychologii

Proces badawczy przebiega w określony, uporządkowany sposób. Można też wyodrębnić w nim pewną liczbę etapów, faz czy ogniw. Jedną z możliwych propozycji podziału procesu badawczego na **etapy** przedstawiłem w: Brzeziński (2011). Nawiązując do tego modelu procesu badawczego w psychologii, skonstruowałem tabelę 1, w której wyszczególniłem – dla każdego etapu (kolumna 1) – zakres czynności badawczych do wykonania (kolumna 2), cel, jaki badacz zamierza osiągnąć (kolumna 3), oraz spodziewane efekty (kolumna 4).

Ocena **istotności zmiennych** (dokładniej: potencjalnej istotności, gdyż ocena dokonywana jest na gruncie aktualnej, niekoniecznie przecież prawdziwej, wiedzy badacza) z punktu widzenia określonej – znajdującej się w centrum zainteresowań badawczych psychologa – zmiennej zależnej jest szczególnie zaakcentowana w dwóch pierwszych etapach procesu badawczego. Przyjrzyjmy się więc tym dwóm etapom.

Punktem wyjścia każdego badania naukowego w naukach empirycznych jest sformułowanie **problemu badawczego** pod postacią **pytania** (chciałoby się napisać – niebanalnego) dotyczącego zmiennej zależnej, a dokładniej jej związków ze zmienną (zmiennymi) niezależną główną. Badacza interesuje nie tylko znalezienie odpowiedzi na pytanie dotyczące tego, **jakie** to zmienne niezależne można uznać za zmienne istotne (dokładniej: zmienne niezależne główne) dla danej zmiennej zależnej Y, ale – czy nawet przede wszystkim – chciałby on znaleźć odpowiedź na pytanie o **kształt związku** łączącego obie zmienne. W naukach behawioralnych zwykło się rozpatrywać dwie klasy związków, w które wchodzi zmienna zależna Y ze zmienną dla niej istotną X: związki **prostoliniowe** i związki **krzywoliniowe.**

Także na gruncie danej wiedzy badacza formułowana jest – jego zdaniem najbardziej prawdopodobna – odpowiedź na pytanie badawcze, czyli **hipoteza.**

Tabela 1. Struktura procesu badawczego w psychologii (według ujęcia Brzezińskiego, 2011)

1	2	3	4
Etap	Zakres czynności badawczych	Cel	Spodziewane efekty
I.	Problem i hipoteza	1. Sformułowanie pytania badawczego o uwarunkowania zmiennej zależnej 2. Sformułowanie hipotezy badawczej traktującej o zależności zmiennej zależnej od danej zmiennej niezależnej (także o interpretację przyczynowo-skutkową zaobserwowanych związków)	1. Wyjaśnienie nowych faktów lub „lepsze” niż dotychczasowe wyjaśnienie danych faktów 2. Powiększenie wiedzy psychologicznej (naukowej) o determinantach ludzkich zachowań
II.	Zmienne	Wyłonienie zmiennych niezależnych, które badacz uznaje, wstępnie, za istotne dla danej, o której traktuje problem badawczy, zmiennej zależnej	1. Zdefiniowanie zmiennej zależnej 2. Ustalenie listy zmiennych niezależnych uznanych przez badacza, wstępnie, za istotne dla zmiennej zależnej 3. Podział zmiennych niezależnych uznanych, „na razie”, za istotne dla zmiennej zależnej na: ważne (główne i uboczne) oraz zakłócające 4. Uznanie, czy rozpatrywana struktura zmiennych ma charakter izolowany czy interakcyjny 5. Zdefiniowanie zmiennych niezależnych uznanych za istotne dla zmiennej zależnej
III.	Operacjonalizacja zmiennych	Projektowanie metod pomiaru zmiennych: zależnej i niezależnych	1. Określenie statusu pomiarowego zmiennych (według klasyfikacji Stevensa) 2. Konstrukcja nowych metod jakościowych (obserwacja, wywiad, analiza wytworów) 3. Konstrukcja nowych testów psychologicznych lub adaptacja kulturowa testów obcojęzycznych 4. Dobór (konstrukcja) aparatury eksperymentalnej i pomiarowej
IV.	Model badawczy	Zaplanowanie, zgodnie z dokonaniem wyborem, modelu badania empirycznego w celu testowania hipotezy – model eksperymentalny vs model korelacyjny	Uzyskanie planu badania gwarantującego wysoką trafność wewnętrzną

V.	Próba	Dobór osób (na jeden z kilku możliwych sposobów) w celu przeprowadzenia zaplanowanego badania empirycznego – dobór losowy vs dobór celowy	<ol style="list-style-type: none"> 1. Przeprowadzenie badania (maksymalizacja trafności wewnętrznej) zgodnie z zaplanowanym modelem 2. Umożliwienie przeprowadzenia zaplanowanych analiz statystycznych 3. Zapewnienie badaniu wysokiej trafności zewnętrznej
VI.	Przeprowadzenie badania	Przeprowadzenie badania według założonego modelu, dokonanie pomiaru zmiennych uwzględnionych w planie badania (zebranie danych)	<ol style="list-style-type: none"> 1. Uzyskanie wyników w warunkach zapewniających wysoką trafność wewnętrzną badania 2. Uzyskanie wyników obciążonych jak najmniejszym błędem zastosowanych narzędzi
VII.	Analiza statystyczna	<ol style="list-style-type: none"> 1. Przeprowadzenie opisu statystycznego uzyskanych wyników 2. Zastosowanie adekwatnych do założonego planu badania statystycznych testów istotności różnic i miar siły związku zachodzącego między zmiennymi 	<ol style="list-style-type: none"> 1. Poznanie rozkładów wyników zmiennej zależnej w porównywanych grupach – eksploracyjna analiza danych 2. Poznanie siły związków zachodzących między analizowanymi zmiennymi 3. Ocena wielkości różnic występujących między porównywanymi grupami pod względem wartości wskaźników analizowanych zmiennych zależnych 4. Ocena wielkości efektów ES
VIII.	Ocena badania i hipotez oraz: sformułowanie i uogólnienie wniosków	<ol style="list-style-type: none"> 1. Ocena poprawności metodologicznej przeprowadzonego badania i – jeżeli zajdzie taka konieczność – dokonanie niezbędnych korekt i uzupełnień – określenie trafności wewnętrznej badania 2. Ocena zakłócającego wpływu osoby badacza i osoby uczestniczącej w badaniu na wyniki przeprowadzonego badania 3. Ocena etyczności badania 4. Odniesienie się do hipotez w świetle rezultatów zastosowanych procedur statystycznych 5. Sformułowanie wniosków 6. Uogólnienie wniosków z poziomu próby na poziom populacji – określenie trafności zewnętrznej badania 	<ol style="list-style-type: none"> 1. Akceptacja (w sensie: aprobowanego przez badacza jej uprawdopodobnienia) hipotezy albo jej odrzucenie 2. Poznanie rzeczywistej trafności zewnętrznej badania – precyzyjne określenie zakresu uogólniania wniosków odnośnie do uzyskanego rezultatu badawczego

Źródło: tabela 4.4 (Brzeziński, Zakrzewska, 2010, s. 206) i tabela 3 (Brzeziński, 2010, s. 176-177)

Badacz, mimo że jest zainteresowany tylko określeniem zależności zmiennej Y od jakiejś określonej zmiennej X , to nie może zignorować tego, że obok tej wyróżnionej zmiennej niezależnej (głównej) na Y oddziałują także inne ważne zmienne. Powinien tedy odtworzyć, z jak największą wiernością, zbiór zmiennych istotnych dla zmiennej zależnej Y (ważnych, a także zakłócających). Badacz – na gruncie swojej wiedzy (podkreślamy, że mamy na myśli jedynie wiedzę naukową, racjonalną) – tworzy zbiór zmiennych niezależnych, uznawanych przez niego za zmienne istotne dla zmiennej Y . Zbiór ten może (i na ogół tak jest) różnić się od zbioru zmiennych *de facto* istotnych dla Y – w szczególności część zmiennych może być rzeczywiście istotna dla Y , a część z nich badacz mógł, na gruncie swojej wiedzy (wszak niepełnej), mylnie uznać za zmienne istotne dla Y . Będziemy więc odróżniać zbiór zmiennych *de facto* istotnych dla Y od zbioru zmiennych uznawanych przez badacza, na gruncie jego wiedzy, za zmienne istotne dla Y . Ten pierwszy będziemy nazywać **przestrzenią zmiennych istotnych dla Y** – w skrócie: P_Y . Z kolei ten drugi określać będziemy mianem **obrazu przestrzeni zmiennych istotnych dla Y** – w skrócie: $O(P_Y)$. Na przykład, jeżeli zmienną zależną Y są „osiągnięcia szkolne”, to do P_Y należy, między innymi, „IQ”, a badacz może, mylnie, do $O(P_Y)$ zaliczyć zmienną niezależną „narodowość ucznia”.

Trafne rozpoznanie zbioru zmiennych istotnych dla Y umożliwi badaczowi poddanie ich efektywnej kontroli i określenie procentowego udziału wprowadzonych przez nie wariancji cząstkowych zmiennej Y w całkowitej wariancji zmiennej Y . Zauważmy jeszcze, że badacze, zajmujący się tym samym problemem badawczym, mogą się różnić, jeśli chodzi o budowane przez nich obrazy przestrzeni zmiennych istotnych dla Y . Także ten sam badacz może, przy kolejnych próbach rozwiązywania tego samego problemu badawczego, dokonywać modyfikacji „wyjściowego” $O(P_Y)$, tak że każdy następny, tworzony przez niego $O(P_Y)$, będzie wierniejszą rekonstrukcją P_Y . Można zatem postęp badań przedstawić pod postacią łańcucha, którego ogniwami są $O(P_Y)$ o coraz to większej trafności.

Badacz nie poprzestaje jedynie na próbie odtworzenia P_Y , ale próbuje też ustalić **porządek istotnościowy** w obrębie zmiennych zaliczonych przez niego do $O(P_Y)$. Porządkuje tedy zmienne uznane za istotne od najbardziej do najmniej istotnych dla Y , tworząc jej strukturę (istotnościową). Rzeczywistej **strukturze przestrzeni zmiennych istotnych dla Y** (w skrócie: S_Y) odpowiadać więc będą jej **obrazy** tworzone przez różnych badaczy podejmujących ten sam problem badawczy (w skrócie: $O(S_Y)$).

Wreszcie zmienne mogą oddziaływać na Y w izolacji albo wchodząc ze sobą w interakcje. W pierwszym przypadku będziemy mówić o **izolowanym rodzaju** S_Y , zaś w drugim – o **interakcyjnym rodzaju** S_Y . Badacz, próbując rozpoznać **rodzaj** (izolowany vs interakcyjny) struktury przestrzeni zmiennych istotnych dla Y, tworzy jego – trafny albo nietrafny – obraz. Możemy tedy mówić o **izolowanym obrazie...** i o **interakcyjnym obrazie...** Rozliczne badania pokazują, że rzeczywistość, którą bada psycholog, ma raczej charakter interakcyjny aniżeli izolowany. Przykładowo, istotna dla kształtowania się „IQ” jest właśnie **interakcja** dwóch zmiennych – „środowiska” i „genotypu”. Zatem nietrafne byłoby utworzenie, w tym przypadku, $O_{iz}(S_Y)$. To, czy badacz będzie rozpatrywał izolowany $O_{iz}(S_Y)$ czy interakcyjny $O_{in}(S_Y)$, ma znaczący wpływ na wybór przez niego i modelu badawczego, i – co w kontekście tematu niniejszej dyskusji szczególnie nas interesuje – **modelu statystycznego**.

Ten drugi etap procesu badawczego nosi nazwę **etapu istotnościowego**. To, co badacz ustali na tym etapie, będzie bardzo brzemienne w skutki dla sprawdzanej hipotezy (teorii). Efektywnej kontroli będą podlegały przecież tylko te zmienne, które badacz uzna – trafnie czy nietrafnie – za istotne dla Y. Podobnie, jeżeli badacz zdecyduje się na wybór jednego z rodzajów $O(S_Y)$: $O_{iz}(S_Y)$ czy $O_{in}(S_Y)$, to będzie to miało konsekwencje metodologiczne dla wyboru określonego modelu badawczego i statystycznego, zgodnie z którym będzie przebiegało badanie empiryczne i, podkreślmy, statystyczne opracowanie danych empirycznych, w świetle których psycholog albo zaakceptuje hipotezę badawczą, albo uzna ją za nietrafną.

Nie wszystkie modele statystyczne (wyłaniane w etapie 6. procesu badawczego) nadają się do testowania hipotez o wpływie **interakcji** dwóch (i większej liczby zmiennych) na Y. Takimi adekwatnymi do treści owych hipotez modelami statystycznym są:

– model **analizy wariancji** (w dwóch odmianach: jedno- (*univariate*) i wielozmiennowej (*multivariate*): ANOVA i MANOVA (np. Kirk, 1995; Winer, Brown, Michels, 1991; Brzeziński, 2008);

– model **wielokrotnej regresji/korelacji** (np. Cohen i in., 2003; Pedhazur, 1997) umożliwiające testowanie hipotez o istotności zarówno pojedynczych zmiennych, jak i ich interakcji (por. zwłaszcza: Aiken, West, 1991).

4. *Wtórność AS względem „kontekstu odkrycia”*

Dwa pierwsze etapy procesu badawczego mają najbardziej **twórczy**, odkrywczy charakter. To, jakie zmienne badacz uzna za powiązane ze sobą i jaki jest charakter owych powiązań, będzie stanowiło o **niebanalności** danego badania. Można powiedzieć, że na tym etapie badania naukowego badacz artykułuje oryginalne, twórcze pomysły, które mogą zmienić oblicze teoretyczne danego działu psychologii, albo też powieła, z małymi modyfikacjami, cudze pomysły czy formułuje hipotezy banalne w ich treści.

Gdyby się odwołać do wprowadzonego przez filozofów nauki (przez: Reichenbach, 1938/1989) podzielenia procesu poznawczego na dwa konteksty – **„kontekst odkrycia”** (*context of discovery*) i **„kontekst uzasadnienia”** (*context of justification*) – to dwa pierwsze etapy procesu badawczego utożsamiać by można z pierwszym kontekstem, a sześć pozostałych z drugim. Ten pierwszy z wyróżnionych kontekstów wymaga „czegoś” więcej od badacza aniżeli tylko dobrego przygotowania warsztatowego, wymaga kreatywności (owej „iskry bożej”). Z kolei drugi wymaga solidnego przygotowania metodologicznego i umiejętności prowadzenia badań empirycznych, ale także – co powinno zwrócić naszą szczególną uwagę – umiejętności planowania i przeprowadzania AS. W pierwszym przypadku jest się, przede wszystkim, twórcą („architektem” – jak powiedział Arystoteles, 1983, s. 4-5). W tym drugim zaś przypadku jest się, tak naprawdę, biegłym „rzemieślnikiem”. I jedni, i drudzy są potrzebni w nauce, ale o jej postępie decydują, przede wszystkim, ci pierwsi.

Mój pogląd w kwestii miejsca AS w procesie badawczym jest następujący:

AS zawsze jest wtórna względem teorii psychologicznej i względem planu badawczego.

Aczkolwiek konstruując plan badawczy, musimy uwzględnić to, do jakiego modelu statystycznego będziemy chcieli się odwołać. Widać to, na przykład, wyraźnie w eksperymentach odwołujących się do modelu ANOVA. Nawiasem mówiąc, w wielkich monografiach dotyczących planowania eksperymentów (np. Kirk, 1995) w tytule występują słowa: „experimental design”, a treść monografii odnosi się do modelu statystycznego ANOVA. Nie jest jednak tak, że ulegając modzie na jakiś typ AS, „zgodzimy się”, aby jakieś narzędzie statystyczne „dyktowało” zakres eksploracji badawczych. Można, rzecz jasna, wyobrazić sobie taką sytuację, gdy wyniki AS zasugerują nam dalszą pogłębioną analizę struktury zmiennych, z odwołaniem się do nowych analiz itd., ale te nowe analizy będą wtórne, gdyż będą zasugerowane przez pierwotne badanie empiryczne

zagnieżdżone w określonym kontekście teoretycznym. To był dla mnie, przed laty (Brzeziński, 1976), jeden z argumentów za zniesieniem tradycyjnej dychotomii: *kontekst odkrycia vs kontekst uzasadnienia*.

5. Wnioski statystyczne a wnioski badawcze

Dość często (nazbyt często!) artykuły empiryczne, a zwłaszcza te, które mają postać krótkich raportów z przeprowadzonego badania (bywa, że mającego, z punktu widzenia postępu w rozwoju psychologii, charakter „śmieciowy”), kończą się wnioskami, które „zero-jedynkowo” odnoszą się do rezultatu zastosowanej, w ramach procedury NHST, metody oceny hipotezy zerowej.

Ci badacze zdają się jednak zapominać, że nawet test istotności różnic zastosowany w warunkach gwarantujących wysoką **moc testu** (por. przypis 7) informuje badacza jedynie o prawdopodobieństwie (wynoszącym: „ $1 - \beta$ ”), z jakim, w warunkach idealnych, może on odrzucić fałszywą hipotezę zerową. Dobrze, to jest (powinna być) pierwsza część wniosku. Nie trzeba jednak zapominać o drugiej – na tej pierwszej nadbudowanej i wykraczającej poza czyste możliwości interpretacyjne wynikające wprost z modelu zastosowanego testu istotności: różnic czy współczynnika korelacji – części wniosku. Ta zaś musi skupić się na dwóch sprawach.

Pierwsza – to ocena warunków, w których zastosowano test, i tego, czy miały one wpływ na jego wynik. Porównywane próby nie są przecież, tak naprawdę, pobrane według losowania zwrotnego z populacji (a takie jest założenie modelu statystycznego testu istotności różnic – Lehmann, 1968). Czy rzeczywiście badacz kontrolował wszystkie **wpływy zewnętrzne**, które mogły zniekształcać zachowanie się osób badanych w sytuacji badawczej? Oczywiście, że nie. Czy w badaniach o charakterze eksperymentalnym stosował się do **zasady randomizacji**? Poważny badacz na ogół dopełnia tego warunku.

Druga – to ocena **konsekwencji praktycznych** zaakceptowanego przez badacza ryzyka popełnienia błędu I^0 (zawsze w fazie planowania badania, a nie już po jego zakończeniu!). Zbyt rygorystyczny poziom p może sprawić, że „utopimy” ciekawą hipotezę. Z kolei zbyt liberalny poziom p może doprowadzić do upowszechnienia się fałszywego wyniku, który będzie stanowił podstawę dla jakiejś, powiedzmy, procedury terapeutycznej, która może okazać się szkodliwa (np. źle sprawdzony lek, który okaże się toksyczny). Może zatem nie trzymać się sztywno owego „uświęconego” poziomu $p = 0,05$? Może czasami trzeba sięgnąć po $p = 0,001$, a czasem wystarczy $p = 0,10$ (np. w badaniach typu eksploracyjnego)?

Przychyłam się tedy do sugerowanego przez Kinga i Miniuma (2009, s. 25) podziału wniosków na: **wnioski statystyczne i wnioski badawcze**.

III. REPLIKACJE (CZY RZECZYWIŚCIE SĄ NIEZBĘDNE?)

Odpowiem krótko: TAK. Psychologia i psychologowie uniknęliby wielu kłopotów (też ośmieszenia i kompromitacji), gdyby doceniało się badania będące powtórką czyichś badań.

Przyjrzyjmy się bardzo głośnej, zwłaszcza w środowisku psychologii społecznej, sprawie profesora psychologii z Uniwersytetu w Tilburgu, Diederika Stapela. Została ona upubliczniona w ostatnich miesiącach ubiegłego roku. Oto stosunkowo młody (rocznik 1966), ale o już uznanej w świecie pozycji naukowej, psycholog społeczny z Uniwersytetu w Tilburgu, Diederik Stapel, okazał się „zdolnym” oszustem, który wyprowadził w pole nawet redakcje tak renomowanych czasopism, jak *Journal of Personality and Social Psychology* czy *Science*. Oszust, specjalizujący się w problematyce uprzedzeń społecznych i stereotypów, oparł swoje liczne artykuły na całkowicie zmyślonych wynikach. Autor popularnego artykułu: „Diederik Stapel. Kanciarze z tytułami naukowymi”, zamieszczonego w polskiej edycji tygodnika *Newsweek* (Stawiszyński, 2011), uważa, że za to, że oszuci są, w jakimś sensie, bezkarni, odpowiedzialne są redakcje czasopism, które nie są skłonne drukować artykułów zdających sprawę z powtórzonych, a nie oryginalnych badań. Jelite M. Wicherts (2011), komentując na łamach prestiżowego *Nature* owe bulwersujące zdarzenia, napisał: „*To scientists in other fields, not sharing data may seem extraordinary; to psychologists it is sadly common practice*” [wyróżnienie – J. M. B.].

Na początku niniejszego artykułu napisałem, że jedną z cech poznania naukowego jest jego **powtarzalność**, możliwość przeprowadzenia jego **replikacji** oraz jego **intersubiektywność** (w sensie podanym przez Ajdukiewicza, 1983). Gdyby psychologowie przywiązywali taką samą wagę do replikacji badań empirycznych – przed ich opublikowaniem – jak ich koledzy z nauk przyrodniczych (por. Sun, Pan, Wang, 2010), to można by było uniknąć takich kompromitujących sytuacji, jak ta wyżej opisana.

Zresztą – na co wskazał w swoim wnikliwym i krytycznym (na temat sprawy Stapela) artykule pracujący na Uniwersytecie Wrocławskim psycholog społeczny

Jarosław Klebaniuk (2011)⁶ – niełatwo przebić się autorom takich „wtórnych” opracowań (chyba że mają „nośne” w środowisku nazwisko – to moja opinia), których zamiarem byłoby neglizowanie rezultatów czyjejs (już opublikowanej, a więc i kompetentnie zrecenzowanej) pracy empirycznej. W grę wchodzi też angażowanie czasu i środków finansowych w pracę, która i tak nie przyniesie spektakularnych profitów (wysoka liczba punktów, uznanie itp.), a czasem może być odebrana jako chęć „odegrania się” (pytanie: o co mu/im chodzi?) na autorze/autorach oryginalnego studium. W pełni podzielam diagnozę sytuacji w psychologii społecznej (a nie sądzę, aby w innych jej działach było inaczej):

Inną sprawą domagającą się zmian są replikacje opublikowanych badań. Nie są one cenione przez redakcje czasopism, a jednym z kryteriów kwalifikacji do druku jest to, na ile prezentowane wyniki są nowe, oryginalne i przyczyniają się do rozwoju dyscypliny. Siłą rzeczy replikacje wobec tak postawionych kryteriów są mniej wartościowe niż badania realizowane po raz pierwszy. Utrudnia to weryfikowanie wcześniejszych (m.in. Stapelowskich) rezultatów, choć jest to przecież jeden z podstawowych postulatów metody naukowej. Nie widzę prostego rozwiązania tego problemu, jednak postęp rozumiany jako weryfikacja nowych hipotez nie może być jedynym celem nauki. Także dotychczasowe ustalenia, zwłaszcza gdy, co jest raczej normą niż wyjątkiem, mają słaby – w sensie siły związku lub proporcji wariacji wyjaśnionej – charakter, wymagają dalszego drażenia. Droga do ustaleń psychologii nie jest prosta i choć trudno się z nadmiernie zagmatwanym, niezbyt wyraźnym ich obrazem przebić, to może warto czasem poprzestać na małym. Diederik Stapel „odkrywał” zależności klarowne i liczne. Teraz wiele z nich trzeba będzie poddać w wątpliwość lub wręcz odwołać. **Może warto dać szansę mniej udanym badaniom.** Prawda w psychologii społecznej nie musi być piękna” (Klebaniuk, 2011) [wyróżnienie – J. M. B.].

W polskiej literaturze psychologicznej znaczącą – i jak na razie – jedyną pracę poświęconą replikacji (także autoreplikacji) opublikował psycholog społeczny Bogdan Wojciszke (2006, 2011)⁷. Warto, aby weszła ona do kanonu literatury obowiązującego na studiach doktoranckich z zakresu psychologii czy szerzej: nauk społecznych.

Publikowanie artykułów na ten sam temat, ale z różnymi modyfikacjami (o które upomniał się w swoim artykule Wojciszke – tamże), odnoszącymi się do modyfikacji składu próby i cech sytuacji badawczej (trafność zewnętrzna!) czy kontroli i operacjonalizacji zmiennych (trafność wewnętrzna!), może się przyczynić do wyostrenia obrazu wykrytej w oryginalnym artykule (zakładam, że

⁶ Na ten artykuł zwróciła mi uwagę prof. Maria Lewicka (UW) – której za to dziękuję.

⁷ Czytelnika odsyłam też do interesującej pracy zbiorowej pod redakcją J. W. Neuliepa (1991): *Replication research in the social sciences*.

nie powstał on metodą Stapela) zależności pretendującej do statusu prawa nauki⁸. Chciałbym tu wspomnieć o reanalizach prawa Yerkesa i Dodsona (1908), opisanych w pracy Teigena (1994).

Replikacje badań, a zwłaszcza tych kontrowersyjnych, mają nie tylko swój sens „kontrolny” (realizacja Ajdukiewicza zasady **intersubiektywnej kontrolovaności**), ale gdy będą dobrze poprowadzone od strony AS, to ich wskaźniki ES będą podlegały **kumulacji** w przeprowadzonej metaanalizie. Zatem, raz jeszcze chciałbym podzielić opinię Kinga i Miniuma (2009, s. 357) – tym razem odnośnie do wartości **metaanalizy**:

[...] pomimo pewnych niedoskonałości metaanaliza już pomogła skupić uwagę na wielkości efektu i na wspieraniu rozwoju nauki, który jest rozumiany jako proces kumulatywny (a nie działanie zmierzające do osiągnięcia pewnej magicznej wartości *p*). Nie możemy zapomnieć, że „kolejne próby jej [hipotezy] weryfikacji mogą się okazać albo silniejsze, albo słabsze” (Fisher, 1960, s. 25).

I znowu wróciliśmy do wskaźników *wielkości efektu*. Nie da się od nich uciec!

IV. CZTERY POZIOMY REALIZACJI SEKWENCJI: TEORIA → REZULTAT BADAWCZY

Główną myślą przewodnią niniejszego opracowania jest stwierdzenie, że to teoria (tu: psychologiczna) czyni sensownym cały projekt badawczy i że będąca w centrum naszego zainteresowania AS jest na „usługach” tejże. Jak powiedział znany biolog François Jacob (1973, s. 32): „w wymianie między teorią a doświadczeniem teoria zawsze zaczyna dialog jako pierwsza”.

Relacja: „teoria–empiria” realizowana jest według mnie na czterech poziomach:

Poziom 1. Psycholog poddający empirycznej kontroli jakąś hipotezę musi rozpatrzyć ją w kontekście paradygmatu, z którego wywiedziona została określona teoria stanowiąca ontologiczne i epistemologiczne „tło” dla sprawdzanej hipotezy. Analizując związki zachodzące między sprawdzaną hipotezą a innymi, pokrewnymi teoriami, psycholog musi pozostawać w obrębie danego paradyg-

⁸ Jeśli chodzi o polskie czasopisma psychologiczne, to jedynie redakcja *Psychologii Społecznej* informuje potencjalnych autorów (wręcz zachęca ich!), aby składali artykuły, w których przedstawialiby „nietrywialnie replikacje wyników”.

matu. Odnoszenie hipotezy wyrażonej w języku teorii T^1 , należącej do paradygmatu P^1 , do jakichś elementów teorii T^2 , należącej do paradygmatu P^2 , nie może przynieść wartościowych poznawczo rezultatów, gdyż badacz równocześnie używa dwóch nieprzekładalnych na siebie języków. Mówi „coś”, ale nie bardzo wiadomo, w system jakiej wiedzy to „coś” można wbudować. Nie można bowiem – żeby odwołać się do jakiegoś przykładu – odwoływać się jednocześnie do języka ortodoksyjnej *psychoanalizy* (jeden paradygmat) oraz do języka *behawioryzmu* (drugi paradygmat). Oba bowiem paradygmaty reprezentują całkowicie odmienne punkty widzenia na naturę ludzkiego zachowania i jego determinant.

Także badanie empiryczne musi być zaplanowane z uwzględnieniem kontekstu tego samego paradygmatu. A zatem badacz musi respektować zasadę, którą można nazwać *zasadą niesprzeczności paradygmatycznej* i którą można wyrazić w następujący sposób: zmienne teoretyczne i zachodzące między nimi relacje, o których traktuje hipoteza badawcza, powinny być wyprowadzone z tego samego paradygmatu.

Niestety, młodsze pokolenia psychologów dość często ignorują ową zasadę i formułują, może i ciekawie wyglądające i operujące śmiałymi zestawieniami terminologicznymi, *quasi-teorie* stanowiące zlepek elementów należących do różnych paradygmatów. Co więcej, projektując badanie, wcale nie dbają o powiązanie jego modelu z wyjściowym – dla sprawdzanej hipotezy – paradygmatem.

Poziom 2. Projektując badanie, musimy wpieryw nadać *sens teoretyczny* terminom (zwłaszcza nowym) występującym w hipotezie. Pozostając w obrębie tej samej teorii, nadajemy – uprzednio **zdefiniowanym na gruncie określonej teorii psychologicznej** – terminom (zmiennym) teoretycznym *sens empiryczny*, przeprowadzając zabieg *operacjonalizacji* (jak to się robi, pokazała – wychodząc z założeń idealizacyjnej teorii nauki – Hornowska, 1989). Ten zaś wymaga bądź posłużenia się gotowymi metodami pomiaru interesujących nas zmiennych (np. pomiaru IQ za pomocą WAIS-IV), bądź skonstruowania nowych metod. Konstruując test psychologiczny, musimy też odwołać się do określonej teorii testu. Tylko na gruncie takiej teorii skonstruowany przez badacza test stanowi wartościowe narzędzie. Jakie to teorie wchodzi w grę i w jakich monografiach zostały zaprezentowane, ukazuje tabela 2.

Psycholog, który ukończył studia, na których nie przedstawiono mu rzetelnie problematyki psychometrycznej (a – niestety – jest to możliwe), będzie bezradny. Mało tego, nie będzie wiedział, dlaczego recenzent dobrego czasopisma „się czepia”.

Poziom 3. Uzyskane wyniki, powiedzmy – testu psychologicznego, poddane są analizie statystycznej. Powstaje jednak problem: jaką przyjmą one postać? W zasadzie można mówić o trzech postaciach, jakie mogą przyjąć wyniki (por. tabela 2):

– mogą to być wyniki surowe (np. suma prawidłowych odpowiedzi w jakimś teście dla poszczególnych osób badanych);

– wyniki surowe mogą być **przekształcone** (standaryzowane): tzw. *scaled scores* w Skali Wechslera, steny, teny, tetrony itp.;

– wyniki poszczególnych narzędzi mogą być **zintegrowane** i przedstawione pod postacią profilu psychometrycznego (np. profil Skali Inteligencji) czy *wyniku czynnikowego (factor score)*. Przykładowo, wyniki cząstkowe testów wchodzących do jednego czynnika Skali Wechslera będą tworzyć tzw. wynik czynnikowy (*factor score*) – jak to się robi w odniesieniu do skali WAIS-R(PL), pokazała Zakrzewska (2000).

W zależności od postaci wyniku badacz sięga po mniej lub bardziej zaawansowaną metodę AS. Przykładowo, w odniesieniu do Skal Inteligencji czy wielowymiarowych kwestionariuszy osobowości, których wynik przyjmuje postać **profilu**, najlepszym wyborem jest sięgnięcie po metody *analizy intra- i interprofilowej* (por. odsyłacze bibliograficzne w tabeli 2). Nie zalecam, a wręcz przeciwnie – zniechęcam do analiz porównawczych (dwugrupowych czy wielogrupowych) prowadzonych dla każdego testu Skali Inteligencji czy wymiaru kwestionariusza osobowości odrębnie (np. za pomocą testu *t* Studenta).

Poziom 4. Przeprowadzone analizy porównawcze, z odwołaniem do prostych analiz (np. istotność różnic dwóch średnich za pomocą testu *t* plus wskaźnik ES), albo – dziś już coraz częściej – zaawansowanych, typu MANOVA, MCR plus *analiza ścieżek*, model *równań strukturalnych* (SEM, por. np. Konarski, 2009) itp., stanowią podstawę interpretacji psychologicznej, która – co ważne (respektowanie *zasady niesprzeczności paradygmatycznej*) – powinna być poprowadzona w ramach teorii, w której języku sformułowano hipotezy badawcze. Kontynuując nasz przykład ze Skalami Inteligencji Wechslera, można wskazać bogatą literaturę, której autorzy, też odwołując się do własnego bogatego doświadczenia klinicznego, pokazali, jaki sens psychologiczny nadać uzyskanym przez pacjentów wynikom – por. np. Rapaport (1945); Zimmerman i Woo-Sam (1973), Kowalik (1988), Lichtenberger i Kauffman (2009).

Tabela 2. Cztery poziomy realizacji sekwencji:
TEORIA → REZULTAT

1	2	3
Poziom	Teoria	Rezultat
Poziom 1. Konstruowanie zmiennych i budowanie hipotetycznych relacji między nimi → TEORIA badanej zmiennej	1. Teoria psychologiczna: – Inteligencji – osobowości – temperamentu – depresji	1. Definicje teoretyczne, np. definicja inteligencji, temperamentu, lęku, depresji
Poziom 2. Nadawanie sensu empirycznego analizowanym zmiennym → TEORIA operacjonalizacji badanej zmiennej	2a. Teoria psychometryczna: – model wyniku prawdziwego (<i>True Score Theory</i> – por. Gulliksen, 1950) – teoria odpowiedzi na pozycje testu (<i>Item Response Theory</i> – por. Lord, 1980; Hulin, Drasgow, Parsons, 1983/2006; Hambleton, Swaminathan, Rogers, 1991) – teoria uniwersalizacji (Lord, Novick, 1968) – teoria wyniku generycznego (Cronbach i in., 1972) 2b. Teoria konstrukcji aparatury pomiarowej 2c. Teoria narzędzi nie odwołujących się do psychometrii	2. Narzędzia: – test psychologiczny, np. Skala Inteligencji WAIS-IV, kwestionariusz NEO-FFI, Skala Temperamentu Strelaua itp. – aparatura pomiarowa, np. EEG, fMRI, KT, <i>eye tracer</i> itp. – obserwacja – rozmowa kliniczna – wywiad standaryzowany – analiza wytworów, np. rysunek, pismo itp.
Poziom 3. Interpretowanie ilościowe → TEORIA jako rama interpretacji ilościowej uzyskanego wyniku: prostego lub zintegrowanego	3. Teoria statystyczna – teoria przedziałów ufności CI (por. Kline, 2004) – teoria testowania hipotezy zerowej NHST	3a. Konstrukcja wyniku: – wynik prosty (surowy) – wynik standaryzowany (z) – wynik zintegrowany (np. profil, skupienie) 3b. Rezultat interpretacji statystycznej (ilościowej) dokonanej za pomocą: – przedziały ufności CI – testy statystyczne istotności różnic – analiza profilowa PA (por. np. Tabachnick, Fidel, 2001; Brzeziński, 2011) – analizy regresji/korelacji R/CA (np. Pedhazur, 1997)

		<ul style="list-style-type: none"> – analizy czynnikowej FA (np. Thompson, 2004) – wskaźników <i>wielkości efektu</i> (por. Cohen, 1988; Grissom, Kim, 2005; Fritz, Morris, Richler, 2011) – analizy wyników surowych (por. Tukey, 1977; Hoaglin, Mosteller, Tukey, 2000; Loftus, 1996, 2002; Cumming, Finch, 2005)
<p>Poziom 4. Interpretowanie psychologiczne (jakościowe) → TEORIA jako rama interpretacji statystycznej uzyskanego wyniku</p>	<p>4. Teoria psychologiczna</p> <ul style="list-style-type: none"> – inteligencji – osobowości – temperamentu – depresji itp. 	<p>4. Interpretacja psychologiczna (jakościowa) wyniku, np. kliniczna: interpretacja kliniczna profilu <i>Skali Inteligencji D. Wechslera</i> (dla dzieci i dla dorosłych) – por. np. Zimmerman, Woo-Sam (1973), Kowalik (1988), Lichtenberger, Kauffman (2009).</p>

Źródło: na podstawie tabeli 2 (Brzeziński, 2010, s. 174-175)

V. CZTERY POZIOMY AS

Gdyby zastanowić się nad „głębokością” i „rozległością” przeprowadzanych przez psychologów analiz statystycznych, to można – jak miemam – wyróżnić cztery jej poziomy:

Poziom A. Na tym poziomie badacz wnika w strukturę wyniku pojedynczej osoby: ANOVA dla $n = 1$ (por. Stachowski, Brzeziński, 1982); *analiza intraprofilowa* dla $n = 1$ (por. Brzeziński, 2011, rozdz. 19: „Analiza profili psychometrycznych – porównania intraindywidualne i interindywidualne”). Przykładowo, psycholog dokonuje analizy profilu *Skali Inteligencji Wechslera* danej osoby: budowa przedziałów ufności dla wyników testów tworzących profil, analiza dyspersji profili, analiza odchyień poszczególnych wyników od średniego profilu, ocena istotności różnic między wynikami dwóch testów tworzących profil.

Poziom B. Przejście od wyniku pojedynczej osoby do wyników grupy n osób to poprowadzenie AS „wewnątrz” grup n osób (*within-subjects analysis*); ANOVA i MANOVA; *analiza czynnikowa*; *analiza interprofilowa*, *analiza regresji* (por. Winer, Brown, Michels, 1991; Kirk, 1995; Thompson, 2004; Tabachnick, Fidell, 2001, rozdz. 10; Morrison, 1990; Pedhazur, 1997). Jeżeli sięgniemy po ten sam przykład *Skali Inteligencji Wechslera*, to zastosowanie znajdują bardziej złożone

modele statystyczne: ANOVA z powtarzaniem pomiarów (Brzeziński, Maruszewski, 1978) czy – współcześnie – raczej MANOVA.

Poziom C. Dalsze rozszerzenie AS związane jest z poprowadzeniem porównań: dwu- i p -grupowych – analiza danych między p grupami (analiza „między osobami”, *between-subjects analysis*). Wiąże się to z podejściem NHST (t , ANOVA, MANOVA, testy nieparametryczne) lub ze strategią posłużenia się *przedziałami ufności* (CI, por. Kline, 2004; Tabachnick, Fidell, 2001; Siegel, Castellan, 1988; Cumming, Finch, 2005). Tu też mieszczą się „hybrydowe” analizy „między-wewnątrz” z odwołaniem się do modeli ANOVA i MANOVA.

Poziom D. Ten najwyższy poziom AS obejmuje analizę k badań dotyczących p grup poprowadzoną metodą *metaanalizy* (por. Hunter, Schmidt, 1990; Schmidt, 1995; Schwarzer, 2007; Kleka, 2011). Współczesne przeglądy literatury przedmiotu są już prowadzone z zastosowaniem właśnie metaanalizy. Niestety, w polskich czasopismach psychologicznych są praktycznie nieobecne.

VI. WSTĘPNA OCENA WYNIKÓW – ANALIZA WYKRESÓW I ROZKŁADÓW

Jedną z ważniejszych książek z zakresu AS, z którą przed laty się zapoznałem (w trakcie pobytu na Uniwersytecie w Hamburgu) i która – mimo upływu lat – nic nie straciła na swojej aktualności, była obszerna monografia wybitnego statystyka Johna Tukeya (kto z nas nie zna testu HSD jego autorstwa stosowanego w ANOVA?) pt. *Exploratory data analysis* (Tukey, 1977). Wielokrotnie po nią sięgałem. Szkoda, że nie przełożono jej na język polski. Bardzo by się przydała w dydaktyce uniwersyteckiej. Czego się z niej nauczyłem? Ano też i tego, że znacznie ważniejsze od myślenia o zebranych zbiorze wyników jedynie w kategoriach testowania hipotezy zerowej (strategia NHST), a więc o wyborze najlepszego (w sensie: maksymalizującego szanse odrzucenia H_0) testu istotności różnic czy też wyborze najlepszego (też w powyższym sensie) współczynnika korelacji jest dokładne przyjrzenie się wynikom: ich rozkładom oraz rozrzutowi wyników (por. Hoaglin, Mosteller, Tukey, 2000; Loftus, 1996, 2002; Cumming, Finch, 2005). W modelu MRC ważne jest wychwycenie wyników *wpływowych* czy *oddalonych*. Moim zdaniem w pracach psychologicznych (i w nauczaniu AS naszych studentów) zbyt mało wagi przywiązujemy do wnikliwej analizy rozrzutu wyników oraz analiz różnego formatu wykresów wyników. Z niej bowiem często znacznie więcej się dowiemy aniżeli z „uświęconego” $p = 0,05$ (*sacredness of .05*).

VII. NHST VS PRZEDZIAŁY UFNOŚCI

Dość powszechne, żeby nie powiedzieć, że wszechobecne, w praktyce badawczej psychologów jest odwoływanie się do strategii NHST. Celem badaczy jest odrzucenie, za wszelką cenę, hipotezy zerowej. Tym bardziej że polityka redaktorów naczelnych liczących się czasopism naukowych z zakresu psychologii jest taka, że drukują oni jedynie te prace, w których przedstawiono badania prowadzące do odrzucenia H_0 (por. Sterling, 1970; Daniel, 1998; też kontrowersje wokół NHST przedstawione w szerszym kontekście w: Harlow, Mulaik, Steiger, 1997; Maxwell, 2004; Cohen, 1994/2006). Odrzucenie H_0 , niejako automatycznie, usensawnia empirycznie hipotezę alternatywną. Musimy jednak pamiętać, że decyzje statystyczne w strategii NHST są „zero-jedynkowe”. To zaś oznacza, że jeżeli – przykładowo – „nasza” różnica między dwiema średnimi zmiennej zależnej Y w tradycyjnie zaplanowanym i przeprowadzonym eksperymencie nie przewyższy wartości progowej statystyki testu t przy $p = 0,05$, to nie będziemy mogli uznać hipotetycznej zmiennej X za istotną dla analizowanej zmiennej zależnej (w każdym bądź razie artykułu na ten temat nie opublikujemy w szanującym się czasopiśmie psychologicznym). Podobnie, gdy testujemy hipotezy zerowe dotyczące wartości populacyjnej: średniej, proporcji czy współczynnika korelacji. A przecież to, że „nie wyszło”, wcale nie musi oznaczać, że w rzeczywistości nie zachodzi powiązanie Y z X . Mogło bowiem być i tak, że badanie zostało źle zaplanowane albo że badacz nieskutecznie kontrolował (bądź wcale nie kontrolował) zmienne uboczne i zakłócające. Gdy zaś okaże się, że jednak „wyszło” ($p < 0,05!$), to i tak nic nie wiemy o rzeczywistej wartości populacyjnej testowanej różnicy między średnimi czy wartości populacyjnej średniej.

Bezrefleksyjne posługiwanie się testami istotności sprawia, że każdą, nawet bardzo małą różnicę między grupami można uznać za istotną statystycznie na odpowiednio wysokim poziomie istotności p , gdy tylko próba będzie dostatecznie duża. Pouczające jest to, co zrobili Meehl i Lykken (za: Cohen, 1994/2006, s. 110):

[...] Meehl i Lykken sporządzili tablice kontyngencji dla 15 zmiennych i próby liczącej 57 000 studentów szkół ponadpodstawowych, uwzględniając zawód i wykształcenie ojca i matki, liczbę rodzeństwa, płeć, kolejność urodzeń, plany dotyczące dalszej edukacji, postawę rodziców wobec szkoły, lubienie szkoły, wybór szkoły, plany zawodowe na najbliższe dziesięć lat, preferencje religijne, sposób spędzania wolnego czasu i organizacje szkoły. Wszystkie ze 105 korelacji chi-kwadrat utworzonych przez skombinowanie zmiennych metodą „każda z każdą” okazały się istotne, przy tym 96% spośród nich było istotnych na poziomie $p < 0,000001$ [...].

Ten sam problem podniósł William L. Hays (1973, s. 424), który jeden z paragrafów swojego obszernego podręcznika statystyki zatytułował: „Can a sample size be too large?” Napisał w nim: „Trivial associations may well show up as significant results when the sample size is very large”.

Do tego problemu z opieraniem wniosków badawczych jedynie na wielkości p wrócimy w następnym punkcie.

Lepsza, dająca bardziej wnikliwy wgląd w rzeczywistość, jest strategia **przedziałów ufności**, która doprowadza badacza do wyznaczenia przedziału wartości o określonej szerokości, który obejmuje szacowaną wartość populacyjną (różnic czy parametrów). Owa szerokość przedziału zależy od poziomu ufności badacza co do tego, że skonstruowany przedział rzeczywiście pokrywa nieznaną wartość populacyjną. I tak, przy szerokim przedziale wartości będziemy mogli z większą pewnością (np. $P = 99,9\%$) orzekać, że dana wartość leży w tym przedziale wartości, ale nasze ustalenia będą mało precyzyjne. Z kolei chcąc uzyskać większą precyzję oszacowania, będziemy musieli skonstruować przedział węższy, ale poziom ufności, że tak jest, będzie niższy ($P = 90\%$). Z pewnością musimy poszukać owego złotego środka, gdy przy odpowiadającej nam precyzji przedziału (co do jego szerokości) uzyskamy zadowolający nas poziom ufności – przy czym niekoniecznie musi to być poziom: $P = 95\%$. Musimy pamiętać, że dokonując szacowania, z określonym poziomem ufności (np. 95%), nieznaną wartość populacyjną, na przykład średniej μ , nie zmieniamy wartości populacyjnej parametru średniej, zmienia się jedynie, z próby na próbę (co do wartości jego granic), przedział ufności.

Znacznie więcej badacz dowie się z przedziału ufności (zobaczy granice, które obejmują szacowaną wartość danego parametru – przy, rzecz jasna, akceptowaniu określonego poziomu ufności) aniżeli z decyzji nakazującej odrzucenie hipotezy zerowej. Nawet gdy w ślad za nią pójdzie obliczenie wartości wskaźnika ES. Dlatego też współcześnie zaleca się, także dla wskaźników ES, aby podawać charakterystyki przedziałów ufności (APA, 2010, s. 34; Wilkinson i Task Force on Statistical Inference, 1999; JARS, 2008; jak to robić i gdzie o tym przeczytać: Hedges, Olkin, 1985; Grissom, Kim, 2005; Thompson, 2007; Sun, Pan, Wang, 2010). Oddajmy jeszcze głos J. Cohenowi, którego dwie przywołane tu prace (por. Cohen, 1990/2006, 1994/2006), obiegiły, bez przesady, cały psychologiczny świat. Odnosząc się krytycznie do strategii NHST, w punkcie „A więc co robić?” (Cohen, 1990/2006, s. 114) pisał: „[...] moja rada dla badaczy jest więc taka, żeby rutynowo podawali oni wielkość efektu w formie przedziałów ufności”. To zaś, że tak niechętnie badacze zamieszczali informacje

o przedziałach ufności⁹, Cohen tłumaczył tym, że przedziały są „tak kłopotliwie szerokie”! I trudno się z nim nie zgodzić.

VIII. *P* VS *ES*

– JAK MIERZYĆ ISTOTNOŚĆ ZMIENNEJ *X* DLA ZMIENNEJ *Y*?

Od wielu już lat psychologowie (i nie jest ich bynajmniej niewielu!) mylą pojęcie **istotności statystycznej**, wyrażane poziomem istotności statystycznej (*p*; dawniej mówiło się o poziomie α) z pojęciem istotności zmiennej *X* dla zmiennej *Y*. Uważają oni bowiem, że wysoki poziom istotności statystycznej (np. $p < 0,00001$) świadczy o równie wysokiej istotności danej zmiennej *X* dla danej zmiennej *Y*. Nic bardziej mylnego. Utożsamiają zatem istotność zmiennej z prawdopodobieństwem popełnienia błędu I⁰ rodzaju (odrzućenie H_0 , gdy jest ona *de facto* prawdziwa). Sam spotkałem się w trakcie opiniowania jakiejś pracy doktorskiej z taką oto sytuacją. Autor (a może autorka) tej pracy posłużył się modelem ANOVA i uzyskał wysokie wartości *p*. W recenzji poprosiłem o wyliczenie wartości wskaźników ω^2 . Odpowiadając na pytania recenzentów, pominął on moją prośbę. Po obronie, w sytuacji już „rozluźnionej”, zapytałem go: a co ze wskaźnikami *omega-kwadrat*? Usłyszałem następującą odpowiedź: „policzyłem, ale wyszły tak niskie wartości, że promotor doradził mi, abym ich nie podawał, bo to osłabiłoby wnioski”. Ja zaś pomyślałem, że „osłabiłoby” spektakularność obrony. Chciałbym też zauważyć, że w badaniach ankietowych (z reguły prowadzonych na bardzo licznych próbach) mamy na ogół do czynienia z wysokimi (lepszymi od $p = 0,001!$) poziomami, tak rozumianej, istotności analizowanych predyktorów.

Jeżeli nie *p*, to co? We współczesnej psychologii (ale nie tylko) **istotność** zmiennej *X* dla *Y*¹⁰ operacjonalizowana jest poprzez różnorakie wskaźniki *ES*. Jest to powszechna, i zalecana (por. Kirk, 2001; także przygotowany na zlecenie APA przez Wilkinsona i jego zespół roboczy raport: Wilkinson i Task Force on Statistical Inference, 1999) praktyka badawcza i publikacyjna.

Od czasu pierwszych prac Jacoba Cohena (por. 1969, 1988, 1990/2006, 1994/2006; też: Kirk, 1996; zob. ten problem w perspektywie historycznej: Hu-

⁹ Cohen pisał te słowa w 1990 roku; od tego czasu nastawienie psychologów uległo zmianie (ale nie w Polsce), m.in. za sprawą raportu Wilkinsona i współautorów oraz standardów publikacyjnych APA (2001, 2010).

¹⁰ Rozumienie jak np. u Hornowskiej (1989).

berty, 2002) wymagania publikacyjne światowych czasopism psychologicznych są takie, że nie będzie publikowany artykuł, w którym jego autor nie przedstawi wartości wskaźników ES, a poprzestanie jedynie na podaniu wartości statystyki testowej oraz poziomu istotności statystycznej p . ES są we współczesnych psychologicznych pracach empirycznych wszechobecne. Niestety, te standardy publikacyjne w zakresie AS, narzucone w pierw przez wyżej cytowany raport zespołu ekspertów koordynowanego przez Lelanda Wilkinsona¹¹, a potem zaakceptowane przez APA, która wprowadziła je do dwóch ostatnich wydań swojego *Publication manual...* (APA, 2001, 2010), nie znalazły oddźwięku w zaleceniach dla autorów, formułowanych przez redaktorów naczelnych najlepszych polskich czasopism psychologicznych (choć znalazły się w polskich opracowaniach podręcznikowych – np. Bedyńska, Brzezicka, 2007; Francuz, Mackiewicz, 2005; Brzeziński, 2008). A szkoda. Szkoda też i dlatego, że **te artykuły skazane są na swoisty „niebyt”, gdyż nie wejdą do różnych współcześnie poprowadzonych metaanaliz**, które „żywią się” danymi o wartościach wskaźników ES. Zwracali już na to uwagę B. M. King i E. W. Minium (2009), którzy pisząc o metaanalizie, podkreślali, iż:

[...] Metaanalizy nie przeprowadza się po przeczytaniu streszczeń albo rozdziałów artykułów badawczych zawierających wnioski. Dokonuje się jej poprzez wyodrębnienie z analizowanych prac danych oraz ich skumulowanie. I w tym tkwi właśnie cały problem. W tekstach przedstawiających wyniki badań rzadko można znaleźć wielkość efektu [...] i trzeba go szacować, często na podstawie nieadekwatnie podanego poziomu istotności (s. 357).

Jak tedy postępować w tradycyjnym modelu NHST? Odwołam się do ulubionego modelu eksperymentalnego ANOVA. Studentom (doktorantom) tak mówię: jeżeli planuje się przeprowadzenie badania eksperymentalnego wykraczającego poza tradycją uświęcone porównanie dwóch grup (eksperymentalnej i kontrolnej) i sięga po zaawansowane plany eksperymentalne ANOVA, to powinno się uwzględnić wykonanie następujących kroków:

(1) tradycyjne „obwążowanie nieznanego terenu”, albo inaczej – typowe **badania eksploracyjne**: testowanie, za pomocą testu F , hipotez zerowych dotyczących *efektów głównych* poszczególnych czynników i w przypadku odrzucenia H_0

¹¹ Może jednak warto o tym wspomnieć, że inspiracją dla prac zespołu Wilkinsona była słynna praca Jacoba Cohena: *Ziemia jest okrągła* ($p < 0,05$), którą udostępniłmi czytelnikowi polskiemu (por. Cohen, 1994/2006). Warto do niej, od czasu do czasu, zaglądać (a także do drugiego, równie słynnego artykułu Cohena (1990/2006): *O tym, czego się nauczyłem (jak dotąd)*.

na danym p (powiedzmy, że 0,05 – zwracam uwagę, że czasami dobrym wyborem – zwłaszcza gdy eksplorujemy naprawdę nieznaną tereny – jest $p = 0,10!$);

(2a) sięgnięcie po jakiś test wielokrotnych porównań typu *a posteriori/post hoc* (np. popularny i mało konserwatywny test HSD Tukeya z poprawką Bonferroniego) albo

(2b) przeprowadzenie *testów a priori/porównań zaplanowanych* (inaczej zwanych *kontrastami*), gdy mamy precyzyjnie sformułowane (co do treści i ich kierunku) hipotezy (wyprowadzone z teorii!) dotyczące tylko określonych porównań, a nie porównań każdej średniej z każdą; takie postępowanie jest **zalecane** chociażby przez Rosenthala, Rosnowa i Rubina (2000), Sosnowskiego (2004) czy przez Wilkinsona i Task Force on Statistical Inference (1999);

(3) gdy stwierdza się (ANOVA dwu- i więcej czynnikowa) wystąpienie *efektów interakcyjnych*, przeprowadzenie analiz *efektów prostych*;

(4) obliczenie wielkości ES (zalecam obliczanie wskaźnika *omega-kwadrat* (ω^2), który daje bardziej wiarygodną ocenę wariancji Y tłumaczonej przez jakiś czynnik, czy interakcję czynników aniżeli wskaźnik *eta-kwadrat* (η^2) uwzględniony w pakiecie SPSS – por. Hays, 1973; Brzeziński, Stachowski, 1984; Brzeziński, 1983, 2008; także: Vaughan, Corballis, 1969; Dodd, Schultz, 1973; dla użytkowników SPSS: por. Field, 2009, pkt 10.5, 12.6).

Rzecz jasna, to tylko ogólny zarys AS odnoszący się do – dziś to też już klasyka! – eksperymentowania odwołującego się do modelu ANOVA. Bardziej złożone – chociaż w sensie ideowym podobne – będzie postępowanie eksperymentalne odwołujące się do modelu MANOVA; tu także analiza powinna kończyć się obliczeniem wielkości ES (w kwestii ES w MANOVA – por. Tatsuoka, 1993). Niestety, w większości prac drukowanych w polskich czasopiśmie i książkach ich autorzy nie wykraczają poza wartości F i p (rzadziej: wartości η^2 gdyż, przypominam, dostępne rutynowo w SPSS!).

Także odwołanie się do testów nieparametrycznych powinno się kończyć obliczeniem wartości ES – np. test Kruskala-Wallisa \rightarrow wskaźnik E^2 , test Friedmana \rightarrow wskaźnik W , test Manna-Whitneya \rightarrow r_g -Glassa, test Wilcoxon \rightarrow rangowy współczynnik korelacji dwuseryjnej dla par dopasowanych r_c itd.

A gdy sięgamy po test t ? Wówczas mamy w dyspozycji wskaźniki (przypadek danych niezależnych): ω^2 , d Cohena, g Hedgesa, Δ Glassa albo współczynnik korelacji punktowo-dwuseryjnej r_{pbi} ¹².

¹² Nawiasem mówiąc, wskaźniki ES: d , g , r są wzajemnie przekształcalne – jeden na drugi (por. Rosenthal, Rosnow, Rubin, 2000, s. 12; także: Durlak, 2009).

ES mają dziś bardzo bogatą literaturę. W dyspozycji badaczy znajduje się co najmniej kilkadziesiąt różnych wskaźników, które nie dają tych samych wartości dla tego samego zbioru danych. Mogę, na początek, polecić dobre, praktyczne opracowania autorstwa Roberta Rosenthala, Ralpha L. Rosnowa i Donalda B. Rubina (Rosenthal, Rosnow, Rubin, 2000; także: Rosnow, Rosenthal, 2003, 2009) oraz Roberta J. Grissoma i Johna J. Kima (2005). W każdym bądź razie, co chciałbym podkreślić, nie można wypuścić w świat pracy, w której nie będą podane wartości ES!

IX. JEDNOZMIENNOWA AS VS WIELOZMIENNOWA AS

Wszystko zaczęło się, jak dobrze wiemy, od analiz jednozmiennowych, jednowymiarowych (w sensie: jedna zmienna zależna), ale już w latach trzydziestych ubiegłego wieku stworzono teoretyczne podstawy **analiz wielozmiennowych, wielowymiarowych** (w sensie: wiele zmiennych zależnych), które dziś stosujemy. Tak jak uzyskano nową jakość po przejściu od modelu testu *t* do modelu ANOVA i związanych z nim testów porównań wielokrotnych (*a priori*/porównania zaplanowane i *a posteriori/post-hoc*), tak uzyskano kolejną nową wartość po przejściu od modelu ANOVA do modelu MANOVA. Przeszkodą w upowszechnieniu podejścia wielozmiennowego (np. MANOVA czy model *korelacji kanonicznej*) były trudności obliczeniowe. Postęp w konstrukcji szybkich komputerów i dostarczenie ich na biurka badaczy, a także rozwój pakietów statystycznych SPSS, SAS czy STATA – to wszystko miało bardzo duży wpływ na upowszechnienie tych zaawansowanych teoretycznie i obliczeniowo metod statystycznych. Dziś każdy student może w swoim notebooku przeprowadzić skomplikowane analizy statystyczne. Problem tylko w tym, aby je zrozumieć. Niewiele jest w polskich czasopismach sensownych (odpowiedzialnych) – które nie pełnią „dekoracyjnej” roli – analiz wielozmiennowych.

Czy mamy z tego zysk poznawczy? Niewątpliwie tak. I to znaczny. Tak jak ANOVA umożliwiła, w jednym badaniu, ocenę wpływu wielu zmiennych niezależnych (a zwłaszcza ich interakcji) na zmienną zależną, tak MANOVA pozwoliła przezwyciężyć kolejne ograniczenie. Oto psycholog może, zamiast kilku eksperymentów ANOVA (dla każdej zmiennej zależnej odrębnie), przeprowadzić jeden eksperyment MANOVA, w którym posłuży się dwiema lub większą liczbą zmiennych zależnych składających się na jedną *syntetyczną* zmienną zależną. MANOVA także zastępuje klasyczny model ANOVA z powtarzaniem pomiarów

zmiennej Y (por. Weinfurt, 1998, s. 269-271; Stevens, 2002, rozdz. 13: „Repeated measures analysis”, s. 492-557). Myślę, że model MANOVA w dużym stopniu zastąpi w projektowaniu eksperymentów model ANOVA, gdyż dostarcza bardziej wnikliwych rezultatów. Podobnie z modelem MRC, który też dostał rozszerzony po stronie zmiennej zależnej do modelu *korelacji kanonicznej*.

Statystyka wielowymiarowa (*multivariate statistics*) to obszerny, stale się powiększający dział statystyki. Różne problemy techniczne związane z tą grupą metod wymagają jednak odrębnego opracowania. Dobrze wprowadzenie w tę problematykę daje przywoływany już podręcznik Tabachnick i Fidell (2001). Także polscy psychologowie zadbali o propagowanie zaawansowanych metod statystycznych, w tym: MANOVA, MRC, analiza dyskryminacyjna, model regresji logistycznej, poświęcając tej tematyce specjalny numer *Psychologii Społecznej* (por. Sosnowski, 2010).

X. KILKA MYŚLI „NA KONIEC”

Po pierwsze, AS jeżeli ma być sensownie przeprowadzona, to musi być podporządkowana celowi pracy (artykuł, monografia, referat, poster), w której autor prezentuje wyniki przeprowadzonych badań empirycznych. Te zaś służą **sprawdzeniu empirycznemu** (pozytywnemu: *konfirmacja* czy negatywnemu: *falsyfikacja*) wymyślonych (tak, wymyślonych!) przez psychologa hipotez.

Po drugie, „sensowność”, od której zacząłem pisanie podsumowania, związana jest przede wszystkim z osadzeniem zmiennych teoretycznych w określonym **kontekście teoretycznym**. To on „podpowiada” badaczowi, jak **zoperacjonalizować zmienne** (pierwsza okazja do odwołania się do kompetencji badacza w zakresie AS) i jaki **model statystyczny** będzie adekwatny do testowanych hipotez (druga okazja do odwołania się do kompetencji badacza w zakresie AS).

Po trzecie, jeżeli sprowadzić treść hipotezy do ustalenia siły oddziaływania (istotności) zmiennej niezależnej na określoną zmienną zależną (przy założeniu określonego „kształtu” owej zależności), to kluczowy w łańcuchu postępowania dowodowego badacza będzie wybór odpowiedniego (z wielu!) wskaźnika **wielkości efektu** (ES). To dzięki tej wartości badanie ma tak naprawdę szansę dołączenia do innych badań przeprowadzonych w świecie dzięki wykonanej przez kogoś **metanalizie**. Skupianie uwagi na poziomie istotności p , a nie na ES może tę szansę zniweczyć.

Po czwarte, zanim jednak sięgniemy po mniej lub bardziej wyrafinowane metody AS (rozpiętość od testu χ^2 po model *równań strukturalnych*), warto sko-

rzystać z rad przywołanego wyżej Johna Tukeya i bardzo wnikliwie przyjrzeć się **rozkładowi wyników**.

Po piąte, ważne jest skupienie się na strategii **przedziałów ufności**.

Po szóste, cechą charakterystyczną badania naukowego jest jego powtarzalność. Dojrzała nauka odwołuje się do procedury **replikacji** badań. Taki zabieg ma też na celu uniknięcie zbyt pochopnego wnioskowania (w badaniu mogły wystąpić niezauważone błędy) oraz eliminowanie ze społeczności badaczy oszustów (por. dyskutowany wyżej przypadek Diederika Stapela).

Po siódme, AS to tylko narzędzie. Jak każde, może być poprawnie wykorzystane i przyczynić się do wzbogacenia wiedzy psychologicznej. Ale może też wyprowadzić badacza na manowce poznania naukowego i być źródłem **artefaktów** (że odwołam się do tytułu znanej pracy: Rosenthal i Rosnow, 2009).

Jako czynni psychologowie, badacze i nauczyciele akademicy powinniśmy dbać nie tylko o nasze wyniki (od nich wszak w dużej mierze zależy nasza kariera naukowa), ale także powinniśmy zadbać o edukację tych, którzy będą w przyszłości kontynuować nasze prace badawcze. Mówiąc inaczej, powinniśmy naszym studentom i doktorantom **przekazywać nasze doświadczenie, formować ich świadomość metodologiczną** oraz – co nie mniej ważne – **świadomość etyczną** (por. Brzeziński, 2012). Pamiętajmy, że kształtujemy ich „sumienie naukowe” (Ajdukiewicz, 1965, s. 175) i że: „efektem złej nauki jest zła etyka” (Rosenthal, 1996). To zaś jest istotnym elementem składowym misji uniwersytetu. Nasuwa się jednak, w aktualnej pogoni za mierzalnymi efektami, pytanie: w jaki sposób mierzyć wielkość tego specyficznego „efektu”?

BIBLIOGRAFIA

- Aiken, L. S., West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Ajdukiewicz, K. (1965). *Logika pragmatyczna*. Warszawa: PWN.
- Ajdukiewicz, K. (1983). *Zagadnienia i kierunki filozofii. Teoria poznania. Metafizyka*. Warszawa: Czytelnik.
- APA, American Psychological Association (2001⁴). *Publication manual of the American Psychological Association*. Washington, DC: Author.
- APA, American Psychological Association (2010⁵). *Publication manual of the American Psychological Association*. Washington, DC: Author.
- Arystoteles (1983). *Metafizyka*. Warszawa: PWN.
- Bedyńska, S., Brzezicka, A. (red.) (2007). *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii*. Warszawa: Wydawnictwo Szkoły Wyższej Psychologii Społecznej „Academica”.

- Brzeziński, J. (1976). Budowa prototeorii idealizacyjnych. W: L. Nowak (red.), *Teoria a rzeczywistość* (s. 135-162). Warszawa–Poznań: PWN, Poznańskie Studia z Filozofii Nauki, t. 1.
- Brzeziński, J. (1982). Protoidealizacyjny model procesu badawczego. Próba konkretyzacji. *Studia Pedagogiczne*, 44, 59-80.
- Brzeziński, J. (1983). Ocena efektu eksperymentalnego w układach eksperymentalnych analizy wariancji. *Przegląd Psychologiczny*, 25, 1, 155-165.
- Brzeziński, J. (2008). *Badania eksperymentalne w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe Scholar.
- Brzeziński, J. (2010). Czy psychologia znajduje się na metodologicznym rozdrożu? W: J. Grad, J. Sójka, A. Zaporowski (red.), *Nauka – kultura – społeczeństwo. Księga jubileuszowa dedykowana Profesor Krystynie Zamiarze* (s. 155-192). Poznań: Wydawnictwo Naukowe UAM.
- Brzeziński, J. (2011⁵). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.
- Brzeziński, J. (2012). Jakich kompetencji badawczych oczekujemy od psychologa? W: H. J. Grzegówska-Klarkowska (red.), *Agresja, socjalizacja, edukacja. Refleksje i inspiracje* (s. 393-419). Warszawa: Wydawnictwo Akademii Pedagogiki Specjalnej.
- Brzeziński, J., Maruszewski, T. (1978). Metoda badania zgodności profili psychometrycznych. *Przegląd Psychologiczny*, 11, 3, 569-587.
- Brzeziński, J., Stachowski, R. (1984²). *Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych*. Warszawa: PWN.
- Brzeziński, J., Zakrzewska, M. (2010). Metodologia. Podstawy metodologiczne i statystyczne prowadzenia badań naukowych w psychologii. W: J. Strelau, D. Doliński (red.), *Psychologia akademicka. Podręcznik* (wyd. 2 poprawione, t. 1, s. 175-302). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Buczowski, P., Nowak, L. (1979). Idealizacja a istotność. Studium przypadku: Marksowska teoria klas społecznych. W: A. Klawiter, L. Nowak (red.), *Odkrycie, abstrakcja, prawda, empiria, historia a idealizacja* (s. 59-86). Warszawa–Poznań: PWN.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum (wyd. II – 1988).
- Cohen, J. (1990/2006). O tym, czego się nauczyłem (jak dotąd). W: J. Brzeziński, J. Siuta (red.), *Metodologiczne i statystyczne problemy psychologii* (s. 75-99). Poznań: Zysk i S-ka Wydawnictwo.
- Cohen, J. (1994/2006). Ziemia jest okrągła ($p < 0,05$). W: J. Brzeziński, J. Siuta (red.), *Metodologiczne i statystyczne problemy psychologii* (s. 100-118). Poznań: Zysk i S-ka Wydawnictwo.
- Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cronbach, L. J., Rajaratnam, L., Nanda, H., Gleser, G. C. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: J. Wiley.
- Cumming, G., Finch, S. (2005). Inference by eye. Confidence intervals and how to read pictures of data. *American Psychologist*, 60, 170-180.
- Dancey, C. P., Reidy, J. R. (2011⁵). *Statistics without maths for psychology*. Edinburgh: Pearson.
- Daniel, L. G. (1998). Statistical significance testing: A historical overview of misuse and misinterpretation with implications for the editorial policies of educational journals. *Research in the Schools*, 5, 23-32.
- Dodd, D. H., Schultz, R. F. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance designs. *Psychological Bulletin*, 79, 391-395.
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34, 917-928.

- Egiert, R. (2000). *Parafrazy idealizacyjne*. Poznań: Wyd. Fundacji Humaniora.
- Ferguson, G. A., Takane, Y. (2003³). *Analiza statystyczna w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.
- Field, A. (2009³). *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage.
- Fisher, R. A. (1960⁷). *The design of experiment*. Edinburgh, Scotland: Oliver & Boyd.
- Francuz, P., Mackiewicz, R. (2005). *Liczy nie wiedzą, skąd pochodzą. Przewodnik po metodologii i statystyce nie tylko dla psychologów*. Lublin: Wydawnictwo KUL.
- Fritz, C. O., Morris, P. E., Richler, J. J. (2011). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, (Aug.) 8, 1-17.
- Gaul, M. (1990). *Idealizacyjne modele poznania naukowego w psychologii*. Warszawa–Poznań: PWN.
- Grissom, R. J., Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: J. Wiley.
- Hambleton, R. K., Swaminathan, H., Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Harlow, L. L., Mulaik, S. A., Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.
- Hays, W. L. (1973²). *Statistics for the social sciences*. New York: Holt, Rinehart & Winston.
- Hedges, L. W., Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Hoaglin, D. C., Mosteller, E., Tukey, J. W. (2000). *Understanding robust and exploratory data analysis*. New York: Wiley.
- Hornowska, E. (1989). *Operacjonalizacja wielkości psychologicznych. Założenia – struktura – konsekwencje*. Wrocław: Ossolineum.
- Huberty, C. J. (2002). A history of effect size indices. *Educational and Psychological Measurement*, 62, 227-240.
- Hulin, Ch. L., Drasgow, F., Parson, C. K. (1983/2006). Wprowadzenie do teorii odpowiedzi na pozycje testu. W: J. Brzeziński (red.), *Trafność i rzetelność testów psychologicznych. Wybór tekstów* (s. 213-271). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Hunter, J. E., Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Jacob, F. (1973). *Historia i dziedziczność*. Warszawa: Państwowy Instytut Wydawniczy.
- JARS, APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). *American Psychologist*, 63, 839-851.
- King, B. M., Minium, E. W. (2009). *Statystyka w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.
- Kirk, R. E. (1995³). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746-759.
- Kirk, R. E. (2001). Promoting good statistical practice: Some suggestions. *Educational and Psychological Measurement*, 61, 213-218.
- Klebaniuk, J. (2011). *Profesor Stapel na dopingiu. O upiększaniu psychologii społecznej. Racjonalista*, z dnia 14 listopada 2011 r.; <http://www.racjonalista.pl/kk.php/s,7532> (6 stycznia 2012).
- Kleka, P. (2011). Statystyczne kryteria przydatności raportu z badań do metaanalizy. W: J. Brzeziński (red.), *Metodologia badań społecznych. Wybór tekstów* (s. 99-114). Poznań: Zys i S-ka Wydawnictwo.
- Kline, R. B. (2004). *Beyond significance testing. Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

- Konarski, R. (2009). *Modele równań strukturalnych. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe PWN.
- Kowalik, S. (1988²). Diagnostyczne i kliniczne właściwości Skali Inteligencji: W-B I/II, WAIS, WAIS-R. W: J. Brzeziński, E. Hornowska (red.), *Skala Inteligencji WAIS-R Wechslera. Polska adaptacja, standaryzacja, normalizacja i wykorzystanie w diagnostyce psychologicznej* (s. 386-462). Warszawa: PWN.
- Krajewski, W. (1998²). *Prawa nauki. Przegląd zagadnień metodologicznych i filozoficznych*. Warszawa: Książka i Wiedza.
- Lehmann, E. L. (1968). *Testowanie hipotez statystycznych*. Warszawa: PWN.
- Lichtenberger, E. O., Kauffman, A. S. (2009). *Essentials of WAIS-IV assessment*. New York: J. Wiley.
- Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*, 5, 161-171.
- Loftus, G. R. (2002³). Analysis, interpretation, and visual presentation of experimental data. W: H. Pashler, J. Wixted (red.), *Stevens' handbook of experimental psychology*, t. 4: *Methodology in experimental psychology* (s. 339-390). New York: Wiley.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9, 147-163.
- Morrison, D. F. (1990). *Wielowymiarowa analiza statystyczna*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Neuliep, J. W. (red.) (1991). *Replication research in the social sciences*. Newbury Park, CA: Sage.
- Nowak, L. (1980). *The structure of idealization. Towards a systematic interpretation of the marxian idea of science*. Dordrecht: D. Reidel.
- Nowakowa, I., Nowak L. (2000). *Idealization X: The richness of idealization*. Amsterdam-Atlanta, GA: Rodopi, Poznań Studies in the Philosophy of the Sciences and the Humanities, t. 69.
- Pedhazur, E. J. (1997³). *Multiple regression in behavioral research. Explanation and research*. Fort Worth, TX: Harcourt Brace College Publishers.
- Rapaport, D. (1945). *Diagnostic psychological testing* (t. 1). Chicago: Yearbook Publ.
- Reichenbach, H. (1938/1989). Trzy zadania epistemologii. *Studia Filozoficzne*, 7-8, 205-212 (polskie tłumacz. § 1: "The three tasks of epistemology", w: *Experience and Prediction*, Chicago: University of Chicago Press).
- Rosenthal, R. (1991). Replication in behavioral research. W: J. W. Neuliep (red.), *Replication research in behavioral sciences* (s. 1-30). Newbury Park, CA: Sage.
- Rosenthal, R. (1996). Nauka a etyka w przeprowadzaniu badań psychologicznych oraz analizowaniu i przedstawianiu ich wyników. *Czasopismo Psychologiczne*, 2, 37-46.
- Rosenthal, R., Rosnow, R. L. (2009). *Artifacts in behavioral research*. New York: Oxford University Press.
- Rosenthal, R., Rosnow, R. L., Rubin, (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge, UK: Cambridge University Press.
- Rosnow, R. L., Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221-237.
- Rosnow, R. L., Rosenthal, R. (2009). Effect sizes. Why, when, and how to use them. *Zeitschrift für Psychologie/Journal of Psychology*, 217, 6-14.
- Schmidt, F. L. (1995). *Co naprawdę oznaczają dane? Wyniki badawcze, metaanaliza i wiedza kumulatywna w psychologii*. *Czasopismo Psychologiczne*, 1, 19-31.

- Schwarzer, G. (2007). *Meta: Meta-Analysis. R package version 0.8-2*; <http://cran.r-project.org/doc/packages/meta.pdf> (5 stycznia 2012).
- Siegel, S., Castellan, N. J. (1988²). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193-216.
- Sosnowski, T. (2004). Analiza kontrastów: między eksploracją a testowaniem hipotez. *Przegląd Psychologiczny*, 47, 4, 367-378.
- Sosnowski, T. (red.) (2010). *Zaawansowane metody statystyczne. Psychologia Społeczna*, 5, 2-3, numer specjalny.
- Stachowski, R., Brzeziński, J. (1982). Analiza wariancji wyników eksperymentalnego badania zmian monotonicznych wymuszonych w zachowaniu pojedynczej osoby badanej. *Przegląd Psychologiczny*, 24,1, 159-171.
- Stawiszyński, T. (2011). *Diederik Stapel. Kanciarze z tytułami naukowymi*; <http://nauka.newsweek.pl/diederik-stapel--kanciarze-z-tytulami-naukowymi,84437,1,1.html> (5 stycznia 2012).
- Sterling, T. D. (1970). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. W: R. E. Henkel, D. E. Morrison (red.), *The significance test controversy* (s. 295-300). London: Butterworths.
- Stevens, J. P. (2002⁴). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Erlbaum.
- Sun, S., Pan, W., Wang, L. L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology*, 102, 989-1004.
- Tabachnick, B. G., Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.
- Tatsuoka, M. (1993). Effect size. W: G. Keren, Ch. Lewis (red.), *A handbook for data analysis in the behavioral sciences. Methodological issues* (s. 461-479). Hillsdale, NJ: Erlbaum.
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4, 525-547.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis. Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 45, 423-432.
- Tukey, J. B. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Vaughan, G. M., Corballis, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA design. *Psychological Bulletin*, 72, 204- 213.
- Weinfurt, K. W. (1998). Multivariate analysis of variance. W: L. G. Grimm, P. R. Arnold (red.), *Reading and understanding multivariate statistics* (s. 245-276). Washington, DC: American Psychological Association.
- Wicherts, J. M. (2011). *Psychology must learn a lesson from fraud case*; <http://www.nature.com/news/psychology-must-learn-a-lesson-from-fraud-case-1.9513#auth-1> (5 stycznia 2012).
- Wieczorkowska, G., Wierziński, J. (2011⁴). *Statystyka: od teorii do praktyki*. Warszawa: Wydawnictwo Naukowe Scholar.
- Wilkinson, L., Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Winer, B. J., Brown, D. R., Michels, K. M. (1991³). *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wojciszke, B. (2006). Systematycznie Modyfikowane Replikacje: logika programu badań empirycznych w psychologii. W: J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów*. Warszawa: Wydawnictwo Naukowe PWN.

- Wojciszke, B. (2011). Systematycznie Modyfikowane Autoreplikacje: logika programu badań empirycznych w psychologii. W: J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów* (s. 44-68). Warszawa: Wydawnictwo Naukowe PWN.
- Yerkes, R. M., Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482; też: <http://psychclassics.yorku.ca/Yerkes/Law/> (6 stycznia 2012).
- Zakrzewska, M. (2000). Trzy różne ilorazy inteligencji: interpretacja polskiej wersji Skali Inteligencji dla Dorosłych [WAIS-R(PL)]. *Czasopismo Psychologiczne*, 6, 159-169.
- Zimmerman, I. L., Woo-Sam, J. M. (1973). *Clinical interpretation of the Wechsler Adult Intelligence Scale*. New York: Grune & Stratton.

WHAT DOES IT MEAN
THAT THE RESULTS OF RESEARCH CONDUCTED BY PSYCHOLOGISTS
ARE SUBJECTED TO STATISTICAL ANALYSIS?

S u m m a r y

Referring to the current methodological awareness of psychologists, drawing on his own long-standing research and teaching experience, and inspired by works such as Cohen (1990-2006, 1994/2006) or Wilkinson & Task Force on Statistical Inference (1999), the author addresses crucial problems involved in the use of methods of statistical analysis of data derived from psychological research. He also draws attention to the possible sources of misconduct as well as to optimal solutions, whose adoption by researchers will minimize error variance and contribute to a significant reduction of artifacts created. The aim of the article is also to draw attention to the possible sources of scientific misconduct. The author focused on the following the issues: the standard view of the research process, the concept of statistical analysis of data in the research process, the necessity of replicating research findings, the NHST vs. confidence interval opposition, the importance of determining effect size, and the usefulness of conducting multivariate analysis.

Key words: research process, methodological awareness, effect size (ES), NHST, confidence intervals, replication, multivariate analysis.