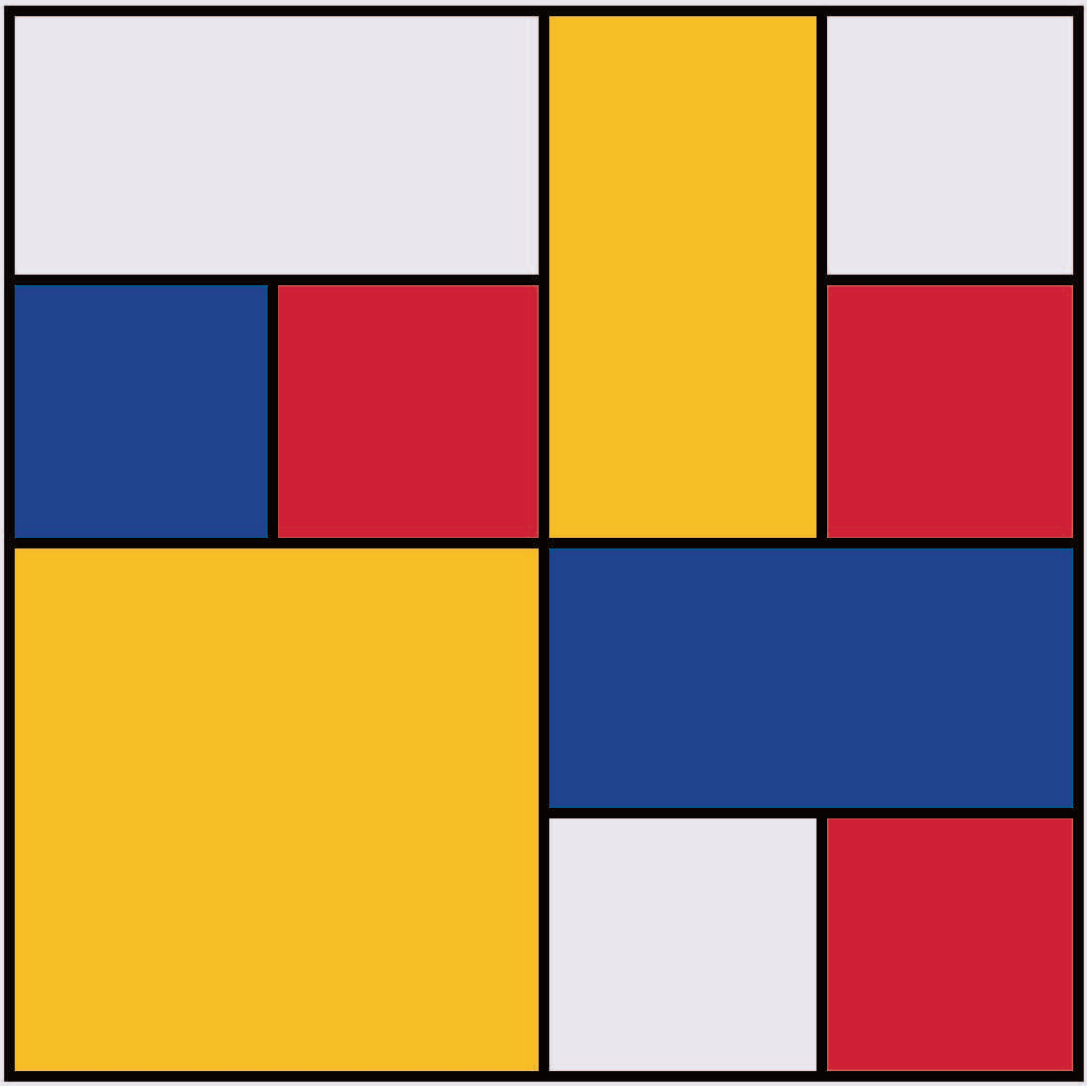
Katarzyna
Grzeszkiewicz-Radulska

Wprowadzenie do analizy wariancji z czynnikami losowymi



Wprowadzenie do analizy wariancji z czynnikami losowymi

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Katarzyna
Grzeszkiewicz-Radulska

Wprowadzenie do analizy wariancji z czynnikami losowymi

WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Łódź 2023

Katarzyna Grzeszkiewicz-Radulska (ORCID: 0000-0002-9155-3588)
– Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny
Katedra Metod i Technik Badań Społecznych, 90-214 Łódź, ul. Rewolucji 1905 r. nr 41/43

RECENZENCI
Jolanta Perek-Białas, Piotr Jabkowski

REDAKTOR INICJUJĄCY
Katarzyna Włodarczyk

REDAKCJA
Weronika Sopala

SKŁAD I ŁAMANIE
Munda – Maciej Torz

KOREKTA TECHNICZNA
Leonora Gralka

PROJEKT OKŁADKI
Polkadot Studio Graficzne Aleksandra Woźniak, Hanna Niemierowicz

https://doi.org/10.18778/8331-176-0

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.09831.20.0.K

Ark. wyd. 7,5; Ark. druk. 10,25

ISBN 978-83-8331-176-0
e-ISBN 978-83-8331-177-7

Wydawnictwo Uniwersytetu Łódzkiego
90-237 Łódź, ul. Matejki 34A
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. 42 635 55 77

Spis treści

Wprowadzenie 7

ROZDZIAŁ I. Role czynników losowych w planach badawczych 13
1.1. Czynniki stałe i czynniki losowe w ujęciu definicyjnym 13
1.2. Czynniki stałe i losowe w perspektywie praktyki badawczej 18
1.2.1. Czynnik losowy „osoby badane” 19
1.2.2. Inne czynniki losowe 23
1.2.3. Plany badawcze bez replikacji i z replikacjami 30
1.2.3.1. Ocena planu 1a 33
1.2.3.2. Ocena planu 1b 39
1.2.3.3. Ocena planu 2b 41
1.2.3.4. Ocena planu 2a 48
1.3. Nieprobabilistyczny dobór próby replikacji 52

ROZDZIAŁ II. Analizy statystyczne 59
2.1. Podstawy analizy wariancji – modele z efektami stałymi w schemacie międzygrupowym 60
2.2. Wyznaczanie oczekiwanych średnich kwadratów $E(MS)$ 72
2.2.1. Źródła zmienności kształtujące wartości średnie dla czynnika stałego w schematach z replikacjami 86
2.3. Statystyka quasi-F 90
2.4. Założenia modeli 93
2.5. Miary wielkości efektu oparte na komponentach wariancyjnych 96
2.6. Moc testów 111
2.7. Przykład analizy danych 119
2.8. Alternatywne strategie wobec modeli z czynnikami losowymi 129

Zakończenie 135

Bibliografia 137

ANEKS 143
Załącznik 1. Moc testów w modelu $T \times r$ 143
Załącznik 2. Moc testów w modelu $r(T)$ 149
Załącznik 3. Dane do przykładu 155
Załącznik 4. Komendy do Edytora Poleceń w IBM® SPSS® Statistics 156

Spis rysunków 159

Spis tabel 161

Spis wykresów 163

Wprowadzenie

W głośnym artykule z 1973 roku Herbert Clark przedstawił intrygującą „zagadkę" metodologiczną. Oto dwóch naukowców równolegle i niezależnie od siebie zajmuje się tym samym problemem badawczym i stosuje tę samą metodę jego rozwiązania. Badają szybkość czytania słów w języku angielskim i obydwaj wysuwają hipotezę, że procesy związane z czytaniem, w tym spostrzeganie, wokalizacja, itd., będą przebiegały szybciej w przypadku rzeczowników niż czasowników. Każdy z badaczy losowo wybiera ze słownika po 10 rzeczowników i czasowników, a następnie dokonuje pomiaru czasu czytania tych 20 słów u każdego z 50 uczestników badania. W rzeczywistości szybkość czytania słów w populacji czasowników jest taka sama, jak w populacji rzeczowników, a więc prawdziwa jest hipoteza zerowa, a nie hipoteza alternatywna, która odpowiada przypuszczeniu obydwu badaczy. Jednocześnie w obu populacjach zróżnicowanie czasu czytania jest duże, co przy małych próbkach oznacza, że mogą one niezbyt wiernie odzwierciedlać rozkład populacji. Jakie są ustalenia badaczy na podstawie dobranych przez nich losowo prób słów i przeprowadzonego testu znaków rangowanych Wilcoxona? Badacz pierwszy otrzymał wynik istotny statystycznie, który był podstawą konkluzji, że przeczytanie rzeczowników zajmuje mniej czasu, a badacz drugi, również na podstawie wyniku istotnego statystycznie, konkludował, że to czasowniki zajmują mniej czasu. Jak to się stało, że wyniki przeprowadzonych testów prowadzą do przeciwstawnych wniosków, a na dodatek żaden z tych wniosków nie odpowiada rzeczywistości?

Zasadniczo uzyskanie wyniku istotnego statystycznie, mimo prawdziwości hipotezy zerowej, nie jest niczym zaskakującym – przyjmując poziom istotności na poziomie np. 0.05 godzimy się na to, że 5 prób na 100 przyniesie wynik uprawniający do odrzucenia hipotezy zerowej, podczas gdy w rzeczywistości jest ona prawdziwa. W tym przypadku sedno problemu leży jednak gdzie indziej. Właściwe rozwiązanie „zagadki" to niedopasowanie metody analitycznej do hipotezy, którą badacze chcieli uprawdopodobnić lub – rozpatrując sprawę z drugiej strony – konkluzja o zbyt szerokim zakresie w stosunku do zastosowanej metody. Badacze wybrali test znaków rangowanych. Wybór tej metody (podobnie zresztą jak wybór jej parametrycznego odpowiednika, czyli testu T dla prób zależnych) wiąże się z uwzględnieniem w modelu tylko jednego czynnika,

w tym przypadku: „część mowy" (czasownik vs. rzeczownik). Metoda ta nie pozwala jednak na uwzględnienie tego, że dane zebrano za pomocą próbki czasowników i rzeczowników, a to powoduje, że odnotowany przez każdego badacza efekt „części mowy" jest *de facto* ograniczony do konkretnych 10 czasowników i 10 rzeczowników użytych w eksperymencie. A skoro wniosek odnosi się do kilku konkretnych rzeczowników i czasowników, to nie powinno dziwić, że jeden ich zestaw może przynieść odmienną konkluzję niż inny zestaw, nawet jeśli oba zestawy zawierają słowa wyłonione losowo z jednego słownika. W modelu zastosowanym przez badaczy wynik istotny statystycznie uprawnia do uogólnienia wniosku na temat efektu czynnika „części mowy" na nieprzebadane jednostki (osoby), ale generalizacja na nieprzebadane czasowniki i rzeczowniki nie jest jeszcze uprawniona. Generalizacja efektu na dwie populacje jednocześnie, tj. populację osób i populację słów, wymaga analiz w oparciu o inny model – taki, który dodatkowo uwzględni czynnik „słowa" i to w charakterze czynnika losowego[1].

Czy przywołanie artykułu, który ukazał się pół wieku temu ma dzisiaj uzasadnienie i czy rozważany w nim przypadek może mieć odpowiedniki w innych obszarach problemowych? Odpowiedź na każde z tych pytań jest twierdząca. Jako przykład rozpatrzmy problem bliski psychologii i socjologii. Badacza interesuje wpływ płci osoby prezentującej jakieś zachowanie. Nie jest możliwe, by uczestnikom eksperymentu zachowanie to zaprezentowała każda kobieta i każdy mężczyzna na świecie – badacz zdecyduje się, by zachowanie to odgrywało przed badanymi co najwyżej kilkanaście kobiet i kilkunastu mężczyzn (być może nawet zdecyduje się na jedną kobietą i jednego mężczyznę). Jeśli badacz nie wprowadzi do modelu losowego czynnika „współpracownik badacza", to odkryty efekt płci dotyczyć będzie jedynie tych konkretnych współpracowników, którzy zostali zaangażowani do eksperymentu. Weźmy przykład badania, którego celem jest porównanie skuteczności różnych metod psychoterapeutycznych. Dopóki model analityczny nie będzie uwzględniał czynnika losowego „psychoterapeuta", to wnioski dotyczące skuteczności poszczególnych metod będą ograniczone do psychoterapeutów, którzy wzięli udział w eksperymencie. Co więcej, nie będzie wiadomo, czy efekt metody jest stabilny i czy charakteryzuje wszystkich psychoterapeutów, a w końcu, jaki udział w objaśnianiu zmian zachodzących u pacjentów ma – obok oddziaływania metody psychoterapii – oddziaływanie psychoterapeuty.

[1] Model analityczny, który Clark (1973) uznał za właściwy jest zaprezentowany w podrozdziale 2.3.

Planowanie eksperymentów w sposób, który drastycznie ogranicza zakres ustalenia badawczego i limituje go do warunków, w których to ustalenie powstało (choć – jak na ironię – pozwala na jego uogólnienie na szerszą populację uczestników) nie jest praktyką optymalną z punktu widzenia rozwoju nauki. Raczej rozwój ten spowalnia i czyni mniej efektywnym, niż niejednokrotnie mógłby być – domaga się przeprowadzenia wielu badań na ten sam temat, tyle że w innych okolicznościach, a potem dokonania metaanalizy, która uczyni przedmiotem badań wyniki, często niespójne, tych „punktowych" eksperymentów. W praktyce zresztą, na skutek braku odpowiedniego materiału, nie zawsze takie metaanalizy udaje się z powodzeniem przeprowadzić, jako że z prostym, mechanicznym replikowaniem eksperymentów rzadko kiedy można się spotkać. Tymczasem trafność zewnętrzną eksperymentu można – i to stosunkowo niewygórowanym kosztem – znacznie podnieść poprzez odpowiednie zaplanowanie eksperymentu i takie skonstruowanie modelu, w którym przynajmniej pewnym okolicznościom, dotąd ograniczającym zakres wniosków, nadany zostanie status czynnika losowego. Taki eksperyment nazywany jest eksperymentem z replikacjami. Wynik eksperymentu z – przykładowo – 20 replikacjami będzie zbieżny z rezultatem metaanalizy przeprowadzonej na 20 eksperymentach bez replikacji (Jackson, 1991: 455).

Książka ta powstała z zamiarem częściowego wypełnienia luki w polskim piśmiennictwie naukowym z zakresu metodologii badań eksperymentalnych. Klasyczne już prace Jerzego Brzezińskiego (2008) oraz Antoniego Sułka (1979, 2002) stanowią fundamentalny wkład w rozwój tej dziedziny. Natomiast podjęta w tej pracy problematyka czynników losowych w planach eksperymentalnych, chociaż dotyczy obszaru znacznie węższego, to jej waga – jak to próbowano wyżej zarysować – wydaje się szczególnie istotna. W polskiej literaturze nie doczekała się ona do tej pory oddzielnego opracowania[2], szczególnie takiego, które byłoby ukierunkowane na specyfikę problemów badawczych w naukach społecznych. Wydaje się ponadto, że sama problematyka czynników losowych jest poniekąd traktowana jako drugoplanowa. Jest w tym racja o tyle, że czynniki losowe prawie nigdy nie grają samodzielnej, ani wiodącej roli w badaniu eksperymentalnym – nie są bezpośrednio ujęte w prawidłowościach, które badacz chce uprawdopodobnić. A jednak mają bardzo duże znaczenie metodologiczne.

[2] W polskiej literaturze zagadnienie czynników losowych w analizie wariancji najszerzej opracował Andrzej Stanisz (2007: 577–611).

W rozdziale I książki staram się te kwestie szczegółowo przedstawić, odwołując się do konkretnych przykładów, analizując różne schematy i praktyki badawcze pod kątem ich ewentualnych niedostatków w zapewnieniu planowi eksperymentalnemu wysokiej trafności. Rozpatruję przy tym problemy badawcze, które odnoszą się do obszaru zainteresowań nauk społecznych. Specyfika zjawisk w tym obszarze wiąże się niejednokrotnie z pewnym, niemającym odpowiednika w naukach przyrodniczych, napięciem na styku bodziec (np. metoda psychoterapii) i nośnik bodźca (np. psychoterapeuta), które wiąże się z tym, że nośnik bodźca nie jest obojętnym elementem układu badawczego i wskazane jest, aby uwzględnić go w modelu jako dodatkowy, najlepiej losowy, czynnik wpływu. Wydaje się, że jako badacze jesteśmy zbyt słabo uczuleni na te problemy – w planach badawczych zwykle rozpatrujemy tylko te czynniki, którymi bezpośrednio manipulujemy i których efekty nas bezpośrednio interesują, traktując inne elementy układu jako niuanse należące do sfery organizacyjnej. Postępując w ten sposób obniżamy nie tylko trafność zewnętrzną ustaleń badawczych, ale także ich trafność wewnętrzną.

Ponadto warto zasygnalizować, że w podejściu do problemu dotyczącego tego, kiedy czynnik może mieć status losowego (czy tylko wtedy, gdy jego poziomy zostały losowo wybrane z populacji poziomów?) zaszła dość duża zmiana. Nastąpiło bowiem przesunięcie ze stanowiska rygorystycznego w kierunku bardziej liberalnego. Jest to efekt burzliwej dyskusji metodologicznej. Niewykluczone, że rozwiązania proponowane przez poszczególne strony nadal będą budzić kontrowersje. Staram się przybliżyć argumenty wymieniane w tej dyskusji, sama zajmując stanowisko bliższe podejściu liberalnemu.

Celem rozdziału II jest pokazanie procedur analitycznych związanych z przeprowadzeniem wnioskowania statystycznego na dwie lub więcej populacje jednocześnie. Przybliżając analizy dla modeli zawierających czynniki losowe skupiam się w głównej mierze na procedurach testowania efektów. Jako uzupełnienie dla wnioskowania statystycznego przedstawiam też procedury szacowania wielkości efektów oraz wyznaczania mocy testów.

W opracowaniu tym ograniczam się do prezentacji procedur dotyczących metody analizy wariancji (ANOVA), która do opracowania danych zebranych w eksperymencie, szczególnie w przypadku grup równolicznych, będzie metodą „pierwszego wyboru". Zdecydowana większość rozważań dotyczy schematów międzygrupowych uzyskanych w planach kompletnie zrandomizowanych. Jednocześnie największą

uwagę poświęcam modelom z efektami mieszanymi, a więc takim, w których obecne są zarówno czynniki stałe, jak i losowe. W przypadku wielu problemów badawczych podejmowanych w naukach społecznych to właśnie te modele będą poprawną alternatywą dla modeli uwzględniających jedynie czynniki stałe.

Niewątpliwie obecność czynników losowych w modelu czyni analizy bardziej złożonymi. Starałam się, by nadmiernie nie komplikować wywodu, a także, by przyjąć bardziej perspektywę badacza niż teoretyka. Pomimo to należy zdać sobie sprawę, że – na tym poziomie zaawansowania analiz – od żadnego oprogramowania statystycznego nie można oczekiwać, że stanie się dla badacza przewodnikiem. Będzie można od niego co najwyżej oczekiwać wykonania obliczeń według skonkretyzowanych przez badacza poleceń. Konieczność przejęcia przez badacza kontroli nad analizami wynika dodatkowo z wielości potencjalnych schematów eksperymentalnych i potrzeby dopasowania działań analitycznych do konkretnego schematu. Staraniem moim było, aby – zachowując standardy pracy akademickiej i wymóg ścisłości wypowiedzi – napisać książkę przystępną i użyteczną dla praktyków.

Chciałbym serdecznie podziękować recenzentom tej książki dr hab. Jolancie Perek-Białas, prof. UJ oraz dr hab. Piotrowi Jabkowskiemu, prof. UAM za wnikliwą lekturę i cenne uwagi. Podziękowania za życzliwą współpracę kieruję do zespołu Wydawnictwa Uniwersytetu Łódzkiego.

Praca powstała w ramach projektu finansowanego przez Narodowe Centrum Nauki UMO-2014/15/B/HS6/01386.

ROZDZIAŁ I

Role czynników losowych w planach badawczych

1.1. Czynniki stałe i czynniki losowe w ujęciu definicyjnym

Czynnik stały (ang. *fixed factor*) to taka cecha, której włączone do badania warianty są ważne dla badacza z teoretycznego punktu widzenia. Zwykle chodzi o przypadek, w którym do badania włączone są wszystkie warianty (poziomy), jakie dane cecha posiada, choć nie musi być to regułą. Kiedy rozpatrywane w badaniu poziomy zmiennej będą stanowić tylko wyimek jej wariantów, ale będą starannie i z rozmysłem wybrane po to, aby utworzyć poznawczo interesujący kontrast (lub kontrasty), wtedy również dany czynnik należy uznać za stały (Jackson, Brashers, 1994a: 1; Lindman, 1992: 129).

Jako przykład może posłużyć czynnik „poddanie terapii metodą X" o poziomach „tak" i „nie"; czy „wielkość dawki leku Y" o poziomach: „wysoka", „niska", „brak leku". Nieco innym przykładem będzie „kierunek studiów", jako że poziomów tego czynnika może być kilkadziesiąt. Niemniej jeśli badacz ograniczy swoje zainteresowania – uzasadniając to odpowiednio – do kilku konkretnych kierunków i do tych celowo wybranych będzie formułować hipotezy oraz dokonywać między nimi porównań, to czynnik „kierunek studiów" należy potraktować jako czynnik stały. Na podobnej zasadzie należy podejść do czynnika „płeć" (czy „tożsamość płciowa"). Porównywanie kobiet i mężczyzn (nieuwzględnienie osób niebinarnych), podobnie jak porównywanie osób niebinarnych i mężczyzn (nieuwzględnienie kobiet), czy porównanie tych trzech kategorii osób, to wszystko będą przykłady użycia w badaniu czynnika stałego (nie zajmuję się tu kwestią uzasadnienia dla stawianych hipotez i tworzenia takich a nie innych porównań).

W charakterystyce czynnika stałego istotna jest także kwestia generalizacji wniosków, czyli możliwości dokonywania uogólnień poza poziomy czynnika, które zostały włączone do badania. W sytuacjach, w których przebadano wszystkie poziomy, jakie czynnik może przyjąć (przykład z poddaniem terapii metodą X), kwestia jest oczywiście bezprzedmiotowa – poziomy czynnika uwzględnione w badaniu odpowiadają populacji jego poziomów, a zatem nie ma już więcej żadnych

poziomów, na które można by rozszerzyć wnioski. Natomiast wtedy, gdy w badaniu znajduje się tylko część wszystkich możliwych poziomów czynnika (przykład z kierunkami studiów), to nadanie mu statusu czynnika stałego oznacza, że badacz nie zamierza rozszerzyć wniosków na poziomy nieprzebadane, a gdyby ewentualnie próbował przeprowadzić takie uogólnienie, to działanie to należałoby uznać za nieuprawnione (Jackson, Jacobs, 1983). Krótko mówiąc, poziomy czynnika stałego tworzą *interesującą badacza populację poziomów*, czy inaczej to ujmując, zbiór poziomów wyczerpujący jego zapotrzebowanie informacyjne.

Czytelnicy artykułów naukowych, w których posłużono się analizą wariancji stykają się z czynnikami stałymi niemal w każdym relacjonowanym badaniu, choć o czynniku stałym pisze się tam zwykle po prostu „czynnik", a dookreślenie „stały" jest jedynie domyślne. W podobnej sytuacji znajdą się czytelnicy podręczników statystycznych oferujących wprowadzenie do analizy wariancji (przykładowo: Wiktorowicz, Grzelak, Grzeszkiewicz-Radulska 2020), czy użytkownicy pakietów statystycznych korzystający z podstawowych procedur tej metody analizy. Z pojęciem „czynnik stały" zetknięcie następuje wówczas, gdy wywód wykładu lub funkcjonalność programu przewiduje inny rodzaj czynnika, a mianowicie „czynnik losowy".

Czynnik losowy (ang. *random factor*) jest cechą, której poziomy włączone do badania zawsze są tylko częścią zbioru wszystkich możliwych jej poziomów. W odróżnieniu od poziomów czynnika stałego, poziomy czynnika losowego nie będą wykorzystywane do tworzenia kontrastów – hipotezy nie będą głosić porównań między poziomami czynnika losowego. Poziomy te bowiem nie przedstawiają sobą ważnych dla teorii odmian danej cechy, a nawet więcej – są wzajemnie zastępowalne, każdy poziom jest tak samo dobry i użyteczny jak inny, w tym sensie nie ma znaczenia, że próbkę poziomów tworzą te, a nie inne poziomy czynnika. Daną próbkę poziomów czynnika losowego można wymienić na inny zestaw poziomów bez konieczności modyfikacji pytania problemowego. Oznacza to także, że gdyby eksperyment miał być replikowany, to w takim powtórzonym badaniu użyto by innych niż poprzednio poziomów czynnika (Jackson, Brashers, 1994a: 4–7). To kolejna cecha, która odróżnia czynnik losowy od stałego. W przypadku czynnika stałego replikacja badania wiąże się z wykorzystaniem zawsze tych samych poziomów (Kirk, 1968: 51). Gdyby z jakiegoś względu miało dojść do zmiany w obrębie poziomów czynnika stałego, to zmianie musiałoby także ulec pytanie badawcze (Jackson, Brashers, 1994: 4).

Nadanie czynnikowi statusu czynnika losowego wiąże się z intencją rozszerzenia wniosków z badania na nieprzebadane poziomy tego czyn-

nika, czyli na populację jego poziomów. W związku z tym dążeniem pojawia się pytanie o sposób ich wybrania. Jeżeli poziomy te zostaną wyłonione losowo z populacji poziomów, to czynnik bezdyskusyjnie staje się losowy. Jak jednak przekonują Clark (1973), Jackson (1992), czy Jackson i Brashers (1994a), sposób wyboru poziomów czynnika nie powinien być jedynym kryterium decydującym o tym, jaki status powinien otrzymać czynnik. W wielu przypadkach istnieją wskazania, by czynnik *traktować jako* losowy, także wówczas, gdy próba jego poziomów nie jest losowo wybrana. Problem tutaj poruszony był przedmiotem kontrowersji i debat, a stanowisko w tej sprawie ewoluowało – nie bez oporów – od rygorystycznego (na przykład Kirk, 1968: 57) do bardziej liberalnego, w rezultacie czego we współczesnych podręcznikach ze statystyki (na przykład Sahai, Ageel, 2000: 7) losowego doboru nie traktuje się już jako warunku koniecznego do uznania czynnika za losowy. Podobnie liberalne stanowisko prezentują badacze zajmujący się problematyką rzetelności i błędu pomiaru (na przykład Shavelson, Webb, 1991: 11). Sygnalizowane tu zagadnienia będą szerzej omawiane w dalszych fragmentach książki.

Na wstępie należy zauważyć, że pierwszorzędnym czynnikiem losowym, obecnym w każdym badaniu, jest „jednostka badana". Mimo to traktowanie uczestników badania jako *czynnika* spotyka się rzadko w podręcznikach (do wyjątków należą między innymi: Keppel 1982, Keppel, Wickens 2004), a ujęcia takiego nie ułatwia nazewnictwo stosowane do określania poszczególnych schematów analizy wariancji, które rzeczywiście ten czynnik pomija. Pominięcie bierze się stąd, że określenia takie jak jednoczynnikowa analiza wariancji (ang. *one-way ANOVA*, one factor design), czy dwuczynnikowa analiza wariancji (ang. *two-way ANOVA*, two factor design) informują nie tyle o liczbie wszystkich czynników uwzględnionych w schemacie, ile o liczbie czynników pełniących rolę zmiennych wyjaśniających[1].

Tymczasem nie tylko poprawnie, ale i pożytecznie jest myśleć o badanych jednostkach (w badaniach społecznych są to zwykle osoby, np. respondenci) jako o poziomach czynnika losowego (Jackson, Brashers,

[1] Można dodać, że to głównie w schematach międzygrupowych, czyli w takich, w których pomiar na uczestniku dokonywany jest tylko raz, dostrzeżenie osób badanych jako czynnika jest o tyle utrudnione, że zmienności pochodzącej od osób badanych nie daje się oddzielić od zmienności wynikającej z błędów pomiaru – całość tej zmienności określa się mianem wariancji wewnątrzgrupowej (ang. *within cells*), czy mniej formalnie – wariancją nieobjaśnioną lub wariancją błędu. Z kolei w planach wewnątrzgrupowych, czyli takich, w których pomiar na uczestniku dokonywany jest więcej niż raz, obecność czynnika „osoby badane" jest już widoczna – czynnik ten jest wyodrębniony jako samodzielne źródło zmienności, a także występuje w interakcji z innymi czynnikami.

1994a: 2). Weźmy przykład eksperymentu, w którym w charakterze badanych wzięło udział – powiedzmy – 30 osób. W tym przypadku należałoby powiedzieć, że czynnik „osoba badana" przyjął 30 poziomów, a to znaczy, że każdy uczestnik z osobna, oznaczony imieniem i nazwiskiem (np. Anna Kowalska) czy chociażby kolejną cyfrą, stanowi jeden z 30 poziomów tego czynnika. Dostarczona wyżej charakterystyka czynnika losowego doskonale odpowiada temu, czym cechują się badani uczestnicy – w badaniu zawsze występuje ich próba, a nie populacja (chyba, że populacja byłaby bardzo mała i badanie mogłoby mieć charakter wyczerpujący); próbkę tworzą osoby potencjalnie zastępowalne przez inne osoby z tej samej populacji; nie tworzy się kontrastów, by porównać poszczególne osoby badane; badanie prowadzi się z myślą o rozszerzeniu ustaleń badawczych na osoby nieprzebadane. Nie jest więc tak, że jednostki badane postrzegać należy jako elementy mające jakiś inny, „specjalny" status, który różniłby je od reszty czynników. Myślenie o badanych jednostkach jako czynniku losowym nie tylko pomaga zrozumieć, czym on tak właściwie jest, ale także stanowi dobry punkt wyjścia do zobaczenia innych czynników w tej roli. Jednak przydatność ujmowania osób badanych jako czynnika losowego na tym się nie kończy – uwidoczni się ona jeszcze w przedstawionej w rozdziale II metodzie dobierania odpowiedniego składnika błędu przy testowaniu istotności danego efektu, a także przy innych procedurach tam przedstawionych.

Badane jednostki to oczywiście niejedyny losowy czynnik, jaki może wystąpić w planie badawczym. Czynników tych może być więcej i – w przeciwieństwie do czynnika „jednostki badane" – będą one miały status zmiennych wyjaśniających. Wróćmy do przykładu zmiennej „kierunek studiów". Jest to przy okazji jeden z tych przykładów zmiennej, co do której badacz ma wybór, jaki status jej nadać – stały czy losowy. Jak już wspomniano, kierunków studiów jest kilkadziesiąt, trudno zatem, by wszystkie zostały uwzględnione w badaniu, a zmienna przyjęła tak wiele poziomów. Bardziej realistyczne jest włączenie do badania tylko części kierunków studiów. I tę część może badacz różnie wybrać i różnie potraktować. Jeśli jest tak, że pewne kierunki interesują go same w sobie (powiedzmy, że będą to prawo, psychologia, matematyka, biologia i historia), zamierza dokonywać między nimi porównań i jednocześnie ograniczyć wnioski do tych kierunków, to czynnikowi „kierunek studiów" powinien nadać status czynnika stałego. Ale można też wyobrazić sobie inną sytuację. Powiedzmy, że władze uniwersytetu noszą się z zamiarem przeprowadzenia badania obejmującego studentów wszystkich kierunków, ale nie wydadzą zgody na badanie, dopóki nie dostaną przekonującego argumentu, że jest szansa na znalezienie różnic między kierunkami.

Żeby dostarczyć władzom podstawy do decyzji najlepiej byłoby przeprowadzić badanie wstępne na próbce kierunków wybranej losowo, a także uznać czynnik „kierunek studiów" za losowy. To, jakie kierunki znalazłyby się w próbce nie byłoby dla badacza w żaden sposób ważne – gdyby mechanizm losowy wskazał inne kierunki jako wchodzące do próbki, to byłyby one równie dobrym wyborem. Badacz nie zamierza bowiem pytać, czy istnieją różnice między konkretnymi kierunkami znajdującymi się w próbie (nie chce porównywać kierunków), ale jest zainteresowany tym, czy istnieje różnica między jakimikolwiek kierunkami w całej populacji kierunków. Gdyby efekt kierunku studiów okazał się nieistotny statystycznie, oznaczałoby to, że jest prawdopodobne, iż różnic między kierunkami (w populacji kierunków) nie ma. Istotny efekt pozwalałby sądzić, że przynajmniej między niektórymi kierunkami w populacji istnieje różnica i ten wynik stanowiłby przesłankę dla władz uczelni do wszczęcia badania pełnego (przykład za: Lindman, 1992: 127).

W rozpatrywanym wyżej przykładzie czynnik „kierunek studiów" – niezależnie od tego, czy traktowany jako stały, czy losowy – był jedynym (oprócz badanych osób) czynnikiem w analizie. W praktyce badawczej częstszą sytuacją byłaby taka, w której czynnik ten byłby „dodatkiem" w badaniu, a główną rolę odgrywałby jakiś inny czynnik stały. Tym czynnikiem stałym, będącym głównym punktem zainteresowania, mogłaby być – powiedzmy – metoda nauczania (przyjmująca warianty: metoda X i metoda Y). Kierunki studiów odgrywałyby w takim badaniu rolę poniekąd służebną, będąc „tworzywem", na którym te metody nauczania można wypróbowywać czy stosować. Hipoteza badawcza będzie w takiej sytuacji odnosić się do czynnika „metoda" (a nie „kierunek studiów") i głosić prymat jednej metody nad drugą. Kierunki studiów będą natomiast wyznaczać *warunki*, czy też *okoliczności*, w których zachodzi prawidłowość dotycząca metod. Teraz, traktując „kierunek studiów" jako czynnik stały, badacz ogranicza zasięg wniosków do przebadanych kierunków studiów, ale też otwiera sobie drogę do ewentualnych porównań między przebadanymi kierunkami. Przyjmując, że do badania włączone zostały prawo, psychologia, matematyka, biologia i historia, rezultatem takiego podejścia może być wniosek typu: „metoda X jest bardziej skuteczna w porównaniu z metodą Y dla prawa, psychologii, matematyki, biologii i historii", czy wniosek: „metoda X jest bardziej skuteczna w przypadku psychologii, matematyki i biologii, podczas gdy metoda Y w przypadku prawa i historii". Natomiast traktując „kierunek studiów" jako czynnik losowy, wnioski na temat porównywanych metod będą dotyczyły kierunków przebadanych i nieprzebadanych, przy czym możliwości porównań zasadniczo się nie przewiduje.

Rezultatem tego podejścia nie będzie żaden z wcześniejszych wniosków. Badanie mogłoby natomiast zakończyć się konstatacją: „metoda X jest bardziej skuteczna niż metoda Y", albo „brak jest podstaw do stwierdzenia różnic między metodami" (przykład za: Jackson i Brashers, 1994a: 6). Przykład ten unaocznia, że tam gdzie wybór między rodzajem czynnika jest możliwy, badacz powinien kierować się typem konkluzji, którym jest bardziej zainteresowany. Pokazuje jednocześnie, że nie musi być tak, iż wniosek zakresowo szerszy jest zawsze bardziej cenny od węższego, szczególnie z punktu widzenia praktyki.

Przedstawione wyżej przykłady czynników losowych dają wstępny i cząstkowy wgląd w cele, jakie te czynniki mogą spełniać w badaniu. Tymczasem celów tych jest co najmniej kilka i mają one duże znaczenie metodologiczne. Mimo to rola czynników losowych jest słabo dostrzegana. Są to czynniki, o których – z wielu względów – można powiedzieć, że są czynnikami „drugiego planu". Zwykle nie grają wiodącej roli w badaniu – najczęściej nie figurują w głównych hipotezach, nie stanowią impulsu do podjęcia badania. Wagę czynników losowych może też – niechcący i niesłusznie – deprecjonować sama ich charakterystyka, szczególnie w punkcie, w którym mowa o tym, że poziomy czynnika losowego nie przedstawiają ważnych z teoretycznego punktu widzenia stanów rzeczy. W tym kontekście zupełnie uzasadnione stają się pytania o sens wprowadzania takich czynników do układu badawczego i rację zajmowania się nimi. Problemom tym poświęcona jest ta książka. Rozpatrzenie wspomnianych kwestii wymaga bowiem szerszego omówienia – najlepiej na kanwie przykładowego badania, dzięki któremu będzie można pokazać, z jakimi problemami metodologicznymi może mierzyć się badacz konceptualizujący swoje badanie. Ogólnikowa odpowiedź, której można w tej chwili udzielić mówiłaby, że czynniki losowe włącza się do planu badawczego dla zwiększenia trafności ustaleń.

1.2. Czynniki stałe i losowe w perspektywie praktyki badawczej

Odwołajmy się zatem do przykładu badania. Powiedzmy, że badacz chce wziąć pod lupę pewną nowatorską metodę nauczania języka obcego. Metoda nauczania może być oceniana pod wieloma różnymi względami, ale załóżmy, że badacza interesuje skuteczność tej metody, którą chce mierzyć za pomocą liczby punktów zdobywanych za poprawnie udzielone odpowiedzi w teście egzaminacyjnym organizowanym na koniec kursu. Ocena skuteczności danej metody powinna nastąpić w wyniku porównania jej z inną metodą. Gdyby bowiem otrzymać wy-

nik mówiący, że osoby nauczane nowatorską metodą uzyskują, np. średnio 75 punktów na 100 możliwych, to trudno byłoby wiążąco ją ocenić – orzec, czy to dużo czy mało? Jak ten wynik plasuje się na tle innych metod czy chociażby w porównaniu z tradycyjną metodą stosowaną obecnie? Powiedzmy więc, że badacz decyduje się na najprostsze rozwiązanie – porówna nowatorską metodę z metodą tradycyjną. Mówiąc językiem metody analizy (tutaj: metody analizy wariancji) powiemy, że czynnik „metoda" przyjmie w badaniu dwa poziomy: m_1 – metoda nowatorska i m_2 – metoda tradycyjna. Mówiąc natomiast językiem procedury terenowej, powiemy o m_1 i m_2 jako o grupach eksperymentalnych, zabiegach badawczych (ang. *treatments*), warunkach eksperymentalnych, czy manipulacjach eksperymentalnych. Zważywszy na cel badania czynnik „metoda" należy uznać za czynnik stały.

1.2.1. Czynnik losowy „osoby badane"

W następnym kroku postępowania podejmowana będzie decyzja o doborze osób badanych, przy czym rozważamy tu wariant, w którym zamierzeniem badacza jest utworzenie dwóch grup osób: w jednej grupie osoby będą nauczane metodą m_1, w drugiej – metodą m_2. Zatem badacz posłuży się **schematem międzygrupowym**, inaczej schematem dla grup niezależnych (ang. *between-subject design*), co oznacza, że każdy uczestnik zostanie poddany tylko jednemu zabiegowi eksperymentalnemu (zostanie przypisany tylko do jednej grupy, do m_1 albo do m_2). Gdyby badacz miał posłużyć się **schematem wewnątrzgrupowym,** inaczej schematem dla grup zależnych (ang. *within-subject design, repeated measures*) oznaczałoby to, że każdy z uczestników będzie poddany kolejno wszystkim zabiegom eksperymentalnym (znajdzie się w grupie m_1, a potem m_2, lub na odwrót). Specyfika rozważanego tu przykładu badania, w którym jednostki uczą się, czyli nabywają trwałe umiejętności, powoduje, że schemat wewnątrzgrupowy nie jest dobrym rozwiązaniem – oddziaływanie pierwszej metody nauczania będzie utrzymywać się w czasie oddziaływania drugiej metody[2].

Przy schematach międzygrupowych kluczową kwestią jest, aby każdy uczestnik został przypisany do grupy (zabiegu badawczego) w sposób losowy (ang. *random assignment*). **Losowe przypisanie** (jednostek do zabiegów badawczych, co jest równoważne z losowym przypisaniem zabiegów badawczych do jednostek), nazywane też rozlosowaniem, jest jedną z metod określanych szerszym mianem **randomizacji**.

[2] Szerzej o wadach oraz zaletach schematów z powtarzanym pomiarem piszą Keppel i Wickens (2004) oraz Brzeziński (2008).

Przypomnijmy, że randomizacja to posłużenie się procesem losowym w intencjonalny sposób. Termin ten stosuje się zarówno do losowego przypisania, jaki i do losowego doboru próby (ang. *random selection, random sampling*) (Cook, Campbell, 1979: 341)[3]. Zabiegów tych nie należy mylić ze sobą, choć oba mogą być zastosowane w jednym badaniu. Mówiąc w pewnym uproszczeniu, celem obu jest osiągnięcie podobieństwa pod względem struktury (rozkładów cech), ale w przypadku losowego doboru chodzi o to, aby próba była podobna do populacji, natomiast w przypadku losowego przypisania jednostek do grup (podprób) chodzi o to, aby to grupy (podpróby) były do siebie podobne (Shadish, Cook, Campbell, 2002: 248). Dodajmy jeszcze, że wyrównanie rozkładów obejmuje wszystkie możliwe cechy – zarówno te mierzone przez badacza, jak i niemierzone.

Dlaczego tak ważne jest losowe przypisanie? Gdyby badacz nie miał kontroli nad przydzielaniem jednostek do grup i nie zadbał o upodobnienie ich struktury, to po zebraniu danych powstałyby wątpliwość, czy różnica między średnimi wynikami testu końcowego (lub brak tej różnicy) odzwierciedla rzeczywiście efekt metody, czy może wystąpiła w wyniku tzw. selekcji (ang. *selection*) (Shadish, Cook, Campbell, 2002: 56), a więc dlatego, że struktury obu grup nie były wyrównane (np. w jednej grupie znalazło się więcej osób uzdolnionych językowo, średni poziom znajomości języka był wyższy w jednej grupie, itd.). Bez zabiegu randomizacji oddziaływanie metody 1. nałożyłaby się na specyficzną strukturę cech osób w grupie 1., a oddziaływanie metody 2. nałożyłoby się na specyficzną strukturę osób tworzących grupę 2. – oddziaływań tych nie sposób byłoby od siebie oddzielić. W konsekwencji wynik badania byłby zasadniczo pozbawiony wartości jako nieniosący ze sobą konkluzywnej informacji. Problem, o którym tu mowa dotyczy **trafności wewnętrznej** (ang. *internal validity*). Eksperyment wewnętrznie trafny to taki, w którym różnica między grupami pod względem zmiennej zależnej wynika wyłącznie z oddziaływania zabiegu eksperymentalnego. Trafność wewnętrzna uzależniona jest od stopnia, w jakim badacz kontroluje zmienne zakłócające, czyli takie, których oddziaływanie może nałożyć się na oddziaływanie zabiegu eksperymentalnego (Vogt, Johnson, 2016: 208; Brzeziński, 2008: 18; West, 2006: 76). Randomizacja osób badanych ma zasadniczo przeciwdziałać nakładaniu się oddziaływań dwóch czynników – metody (czynnika, którego efekt jest głównym przedmiotem badania i który powinien być wyizolowany) oraz selekcji (Shadish, Cook, Campbell, 2002: 55, 249).

[3] Dodatkowe znaczenia terminu randomizacja przedstawia Boik (2005).

Natomiast odwołując się do rozumowania w oparciu o kanon jedynej różnicy powiemy, że randomizacja osób badanych jest jednym z bardzo ważnych narzędzi pozwalających na eliminowanie konkurencyjnych wyjaśnień mówiących skąd – jeśli nie z oddziaływania manipulacji badawczej (tutaj metody nauczania) – mogła wziąć się różnica między wynikami w grupach m_1 i m_2 (tutaj średnimi liczbami punktów uzyskanych w teście końcowym). Dążeniem badacza jest – w czym losowe przypisanie wydatnie pomaga – by porównywane grupy różniły się tylko manipulacjami eksperymentalnymi (tutaj zastosowaną metodą). Wtedy uzyskałby pewność, że różnica między wynikami w grupach wynika tylko z oddziaływania tych manipulacji.

Z kolei wyrażając powyższe zalety randomizacji w języku statystycznym powiemy, że zapobiega ona wprowadzeniu do układu badawczego błędów systematycznych obciążających wyniki (ang. *systematic errors, biases*). Jednocześnie trzeba pamiętać o tym, że mamy do czynienia z procesem losowym, a zatem nie można oczekiwać, że grupy będą idealnie zbalansowane i będą miały identyczne rozkłady cech – po zabiegu randomizacji podpróby mogą się różnić. Im mniejsze liczebnie będą podpróby, tym różnice mogą być większe. Niemniej – niezależnie od wielkości podprób – zróżnicowanie to będzie miało charakter losowy, a nie systematyczny (Keppel, Wickens, 2004: 9). Mówiąc słowami Kirka, sięgając po zabieg randomizacji, badacz zwiększa zróżnicowanie o charakterze losowym, w celu zminimalizowania efektów systematycznych. Ruch ten jest zdecydowanie opłacalny – jak już wcześniej mówiono, efektów systematycznych, powstałych w wyniku działania zmiennych zakłócających, wyeliminować się nie da lub może to być bardzo utrudnione, natomiast efekty losowe, czyli różnice wynikające z losowej zmienności, są dobrze kontrolowane przez testy istotności statystycznej (Kirk, 1968: 7).

Poświęćmy jeszcze uwagę temu, w jaki sposób przeprowadzić losowe przypisanie. W typowej dla badań eksperymentalnych sytuacji, w której z góry określana jest zarówno liczebność uczestników badania, jak i liczebność każdej grupy eksperymentalnej (z zasady dąży się do ich równoliczności), losowe przypisanie powinno być przeprowadzone według reguły **losowej alokacji** (ang. *random allocation rule*), a nie poprzez nieograniczone losowe przypisanie (ang. *unrestricted random assignment*) (Tchach, Berger, 2005). Najprostszy sposób posłużenia się zasadą losowej alokacji polega na zastosowaniu następującej procedury: jeżeli eksperyment przewiduje t grup (zabiegów badawczych) i s jednostek w każdej grupie, to łączna liczba jednostek biorących udział w badaniu będzie wynosić $t \times s = n$. Jednostki te należy

ponumerować kolejno $1, 2, 3, \ldots n$ i przygotować odpowiadające im karteczki z numerami. Wszystkie karteczki należy wrzucić do kapelusza lub podobnego przedmiotu i wymieszać. Pierwsze s karteczek wyciągniętych z „kapelusza" zawiera numery osób, które zostały przypisane do pierwszej grupy. Drugie s karteczek zawiera numery osób przypisanych do drugiej grupy itd. O przypisaniu na zasadzie losowej alokacji mówimy, gdy każda z możliwych sekwencji przypisań ma takie samo prawdopodobieństwo wystąpienia wynoszące $(s!)^t/(s \times t)!$. Przykładowo w eksperymencie, w którym przewiduje się $t = 2$ grup (grupę A i grupę B) i $s = 3$ jednostki w grupie, liczba możliwych sekwencji przypisań wynosi $(s \times t)!/\,(s!)^t = 20$. Dla 6 jednostek biorących udział w eksperymencie, jedna z sekwencji przyjmie postać AAABBB, druga AABABB, itd. Każda z tych dwudziestu możliwych sekwencji będzie miała to samo prawdopodobieństwo wystąpienia równe 1/20. Jak dowodzi Hader (1973) powyższa definicja *nie* jest równoważna definicji nieograniczonego losowego przypisania, którego cechą jest to, że pierwsza i każda kolejna jednostka ma stałe, takie samo prawdopodobieństwo bycia przypisanym do danej grupy, wynoszące $1/t$. Ilustracją zastosowania tej ostatniej metody rozlosowania będzie przykład, w którym przewiduje się obecność dwóch grup, a o przynależności jednostki do grupy ma decydować rzut monetą. Wtedy rozlosowując przykładowo 20 osób, jest bardzo prawdopodobne, że rozmieszczenie jednostek w grupach nie będzie wynosiło 10:10, a więc, że nie uda się zapewnić grupom równoliczności (Tchach, Berger 2005).

Przydzielanie jednostek do grup na zasadzie losowej alokacji może nastąpić także przy zastosowaniu tabeli liczb losowych. Rozwiązanie to będzie zdecydowanie bardziej praktyczne niż metoda losowania „z kapelusza" wtedy, gdy jednostek dostępnych do badania jest więcej niż potrzebuje badacz lub gdy istnieje populacja, z której badacz ma pobrać losową próbę uczestników eksperymentu, a następnie dokonać ich losowego przyporządkowania do zabiegów badawczych. Jednym zdaniem wtedy, gdy chce w jednym badaniu dokonać obu zabiegów randomizacyjnych: losowania próby oraz losowego przypisania. Sposób postępowania w takim przypadku szczegółowo opisują Cook i Campbell (1979: 352). Tutaj zostanie on bardzo krótko zreferowany, gdyż pobranie próby losowej prostej jest szczegółowo omawiane także w popularnych podręcznikach metodologii badań. Jednostki należące do populacji należy ponumerować (stworzyć operat losowania). Numer nadawany poszczególnym jednostkom powinien mieć tyle cyfr, ile cyfr ma liczba osób w operacie. Przykładowo jeżeli liczba ta wynosi 3570, a więc jest liczbą czterocyfrową, to numer, jakim będziemy oznaczać każdą osobę

w operacie, powinien składać się z czterech cyfr, zaczynając od 0001 a skończywszy na 3570. Przyjmując teraz, że badacz potrzebuje do eksperymentu łącznie 200 osób i przewiduje dwie równoliczne grupy eksperymentalne, to korzystając z tabeli liczb losowych, powinien wylosować 200 osób, a następnie pierwszą setkę wyłonionych przypisać do grupy (podpróby) pierwszej, a drugą setkę przypisać do grupy (podpróby) drugiej. W ten sposób utworzone podpróby będą miały podobne struktury, a dodatkowo każda z nich będzie miała strukturę podobną do struktury populacji – każda będzie dla populacji reprezentatywna.

1.2.2. Inne czynniki losowe

Poza kwestią doboru uczestników kolejna decyzja przed badaczem wiąże się z tym, że zastosowanie takiej czy innej metody nauczania wymaga obecności nauczyciela jako *nośnika bodźca*. Decyzja ta zdecydowanie nie ogranicza się do kwestii czysto organizacyjnych czy technicznych, ale wiąże się z bardzo poważnym problemem metodologicznym. Należy bowiem zauważyć, że wprowadzenie bodźca odbywa się tu za pomocą nośnika, który oprócz tego, że przekazuje bodziec, niesie swoje własne, specyficzne cechy. W tym sensie nauczyciel nie jest narzędziem „przeźroczystym" i dlatego nie musi być elementem neutralnym dla układu badawczego. I tak, nauczyciel – oprócz tego, że będzie uczył stosując konkretną metodę (przekazywał bodziec) – będzie także posiadał wiązkę charakterystycznych dla siebie cech. Jedne z nich mogą być nieistotne z punktu widzenia skuteczności prowadzonej przez niego dydaktyki (np. kolor oczu), podczas gdy inne mogą okazać się istotne (np. sposób ubierania się, poczucie humoru, osobowość, poziom motywacji, itd.). Wskażmy przy okazji tę oczywistość, że nie ma dwóch takich samych nauczycieli, a tym bardziej nie ma jakiegoś jednego, wystandaryzowanego nauczyciela-wzorca, choć akurat ten byłby najbardziej pożądany dla układu eksperymentalnego.

Poza przypadkiem metoda nauczania i nauczyciel, analogiczny problem wystąpi w momencie, gdybyśmy chcieli porównywać skuteczność dwóch lub więcej rodzajów psychoterapii. Badanie takie wymagałoby, co oczywiste, obecności psychoterapeuty, tymczasem każdy psychoterapeuta – oprócz tego, że stosuje daną terapię – jest „jakiś", posiada własne, charakterystyczne cechy, przy czym pewne z nich mogą go uczynić bardziej skutecznym lekarzem w porównaniu do innych, nieposiadających tych cech terapeutów (Martindale, 1978).

Podobnie badając reakcje na pewien typ zachowań (np. dominujące i uległe), należy wziąć pod uwagę, że muszą być one odegrane przez konkretną osobę. „Występu" nie da się oddzielić od cech osoby przed-

stawiającej. Problem w tym, że w odbiorze badanych cechy te mogą współgrać mniej lub bardziej z porównywanymi typami zachowań (przykład za Jackson, Brashers, 1994a).

W metodologicznej refleksji badań eksperymentalnych wskazuje się, że elementem niekoniecznie neutralnym w układzie badawczym może być sam eksperymentator (lub jego współpracownik), który swoim zachowaniem na różne sposoby może nieumyślnie doprowadzić do zniekształcenia efektu manipulacji eksperymentalnej (Rosenthal, Rosnow, 2009; Sułek, 1979). Uwaga na osobie eksperymentatora powinna być skupiona szczególnie wówczas, gdy jego rola wykracza poza bycie pasywnym obserwatorem. Wtedy też istnieje ryzyko, że ustalenia badawcze staną się – przynajmniej do pewnego stopnia – funkcją cech eksperymentatora (Bonge, Schuldt, Harper, 1992).

Problem nie pierwszoplanowych, ale potencjalnie nieneutralnych elementów układu badawczego (takich jak wspomniany wcześniej nauczyciel czy psychoterapeuta) nie jest charakterystyczny jedynie dla badań eksperymentalnych. Na gruncie metodologii badań sondażowych, pośród różnych źródeł błędu pomiaru, wiele uwagi i badań od lat poświęca się osobie ankietera (Groves, 1989; Fowler, Mangione, 1990; Sztabiński, 1997). Jego rola w procesie interrogacji z respondentem sprowadza się zasadniczo do dwóch czynności: odczytywania respondentowi pytań oraz utrwalania jego odpowiedzi (a szukając analogii z badaniem eksperymentalnym: podawania bodźców i odnotowywania reakcji). Mimo tego, że nie są to czynności szczególnie skomplikowane, a ankieterzy przechodzą szkolenia, mające ujednolicić ich zachowanie wobec respondentów, to całkowita standaryzacja oddziaływań ankieterów nie jest możliwa. Właśnie te cechy, którymi ankieterzy różnią się od siebie, nie muszą mieć neutralnego wpływu na odpowiedzi, na których udzielenie zdecyduje się respondent. Część z tych cech nie poddaje się standaryzacji z zasady (płeć, wiek, inne cechy społeczno-demograficzne), inne – poprzez szkolenia – są bardziej podatne na „korektę" i ujednolicenie (sposób zadania pytania, sposób reagowania, gdy pytanie jest niezrozumiałe, itd.).

Powyższe przypadki dotyczyły sytuacji, w której człowiek był dostarczycielem i wykonawcą bodźca rozumianego jako rodzaj działania (aplikował metodę nauczania lub psychoterapię, prezentował pewne zachowanie albo zadawał pytanie, itd.). Tę rolę skrótowo określono jako nośnik bodźca. W innych sytuacjach natomiast człowiek może być wykorzystany jako *nosiciel* pewnej cechy, np. płci, koloru skóry, wieku, itp., która w badaniu eksperymentalnym również będzie pełnić rolę bodźca. Naturalnie takie przypadki również są obciążone omawianym problemem,

jako że nie da się zademonstrować badanej cechy np. „podeszły wiek" w postaci wyizolowanej od reszty cech osoby, która ją prezentuje.

Nie jest też tak, że problem niemożności zaaplikowania bodźca w czystej postaci dotyczy tylko sytuacji badawczych, które wymagają obecności człowieka jako nośnika bądź nosiciela bodźca. Przykładowo badając, które pytania-historyjki – te mające zorganizowaną czy zdezorganizowaną strukturę – są łatwiejsze do zapamiętania, stajemy w istocie przed podobną sytuacją – historyjka musi być wypełniona konkretną treścią, żeby móc jej nadać strukturę zorganizowaną albo zdezorganizowaną. Innymi słowy, przedstawiana badanym historyjka – oprócz tego, że będzie miała jedną albo drugą strukturę – będzie także musiała mieć dodatkowe cechy, opowiadać – dajmy na to – o dniu z życia studenta Tomka, albo o wyprawie górskiej, albo o zwyczajach Jakutów, albo być opowieścią na jeden z wielu innych tematów. Będzie też krótka albo długa, napisana językiem potocznym albo nie, itd. Nabierze więc dodatkowych, specyficznych cech (przykład za Keppel, Wickens, 2004). Analogiczny przykład będzie dotyczyć badania reklam. Jeżeliby badano, które reklamy są bardziej perswazyjne – te, w których przedstawia się zalety reklamowanego produktu poprzez porównanie go z rynkowym odpowiednikiem, czy te, w których takiego porównania się nie stosuje – to takiego badania nie można byłoby przeprowadzić na reklamach, które dałoby się scharakteryzować tylko ze względu na to, czy ich twórca sięga po porównanie, czy nie. Nie da się uciec od tego, że reklama musi mieć swoją treść, podyktowaną chociażby rodzajem reklamowanego przedmiotu (przykład za: Jackson, Brashers, 1994a). I na tym tle rodzi się wątpliwość, czy efekt zaobserwowany w reklamach – przykładowo – pasty do zębów, wystąpiłby w reklamach – przykładowo – kosiarek.

Warto teraz sięgnąć do historii dyskusji metodologicznej wokół omawianego tu problemu, by przyjrzeć się sposobom jego nazywania i zaczerpnąć kolejne przykłady. Impulsem do refleksji metodologicznej nad praktyką badawczą stały się badania psycholingwistyczne. Wymienić tu trzeba Colemana (1964), a także Clarka (1973), którzy upominali się o wyższą świadomość metodologiczną badaczy prowadzących eksperymenty. Mówiąc najogólniej, ich krytyka skupiała się na tym, że badacze nie rozpoznają w pełni konsekwencji wynikających z tego, iż analizując oddziaływanie cech języka (np. różnic między czasownikami i rzeczownikami pod kątem czasu, jaki zajmuje ich przeczytanie), używają konkretnego materiału językowego (tu: próbki konkretnych czasowników i próbki konkretnych rzeczowników). Santa, Miller i Shaw (1979), próbując upowszechnić to krytyczne myślenie na szerszym

gruncie badań eksperymentalnych w psychologii, podkreślali, że tym, co aplikujemy jako bodziec, jest nie tyle bodziec w czystej postaci, ile *materiały stymulacyjne* (ang. *stimulus materials*), czy szukając innych określeń – *zindywidualizowane bodźce* (ang. *individual stimulus*), *itemy bodźca*. Być może w języku polskim poręczne byłyby także określenia *ukonkretnione stymulacje*, czy *ukonkretnione realizacje bodźca*. Należy przez nie rozumieć stymulacje, które przybierają różną postać (korzystając z ostatniego przykładu, byłyby to – powiedzmy – ławka, budynek, ptak, paproć), choć reprezentują tę samą manipulację (tu: rzeczownik). I jak argumentowali, skoro stymulacje się różnią, to mogą wprowadzać dodatkową zmienność (ang. *item variation, stimulus variation*) i prowadzić do przeszacowania badanego efektu (tu: różnicy między rzeczownikami i czasownikami). Podkreślano, że to zagrożenie przez wielu badaczy nie jest rozpoznane, podobnie jak nierozpoznane są warunki uprawnionej generalizacji wniosków poza przebadane przypadki (tu: poza przebadane rzeczowniki i czasowniki).

Pod wpływem krytyki, pod lupę wzięto także praktyki badawcze na polu eksperymentów w dziedzinie komunikacji społecznej. Jackson i Jacobs (1983) pisali – podobnie jak poprzednicy – że u podstaw metodologicznych uchybień wcześniejszych eksperymentów leżało niepełne zrozumienie tego, że tym, co jako badacze chcemy uchwycić, są różnice między abstrakcyjnymi *kategoriami* komunikatów, podczas gdy to, co *de facto* aplikujemy (i jedyne co możemy zaaplikować) to idiosynkratyczne *przypadki* badanych kategorii komunikatów. W języku logiki myśl tę można by wyrazić w ten sposób, że jako badacze chcemy formułować wnioski za pomocą nazw ogólnych, jednak przedmiotem naszych stymulacji eksperymentalnych wcale nie są nazwy ogólne, tylko desygnaty tych nazw, a więc przedmioty, które oprócz tego, że mają cechę wspólną wynikającą z przynależności do zakresu tej nazwy, mają także szereg dodatkowych charakterystyk, którymi mogą się różnić.

W metodologii badań eksperymentalnych problem ten odnosi się do **trafności teoretycznej** (ang. *construct validity*), a więc czynności operacyjnych odpowiadających za przejście od konkretnych zabiegów badawczych do konstruktów teoretycznych (a także od jednostek badanych, pomiarów zmiennej zależnej, warunków badawczych do konstruktów teoretycznych) (Shadish, Cook, Campbell, 2002: 20). I jak piszą przywołani przed chwilą autorzy, operacjonalizacje rzadko dostarczają czystej reprezentacji konstruktu, dlatego jedna operacjonalizacja zwykle nie może być jego dobrą reprezentacją – nie tylko nie pokrywa w pełni konstruktu, ale także wnosi swoje własne cechy nie należące do konstruktu.

Na przykład jedna operacjonalizacja konstruktu teoretycznego „mężczyzna", a więc konkretna osoba płci męskiej, posiada swoje indywidualne cechy (jest w określonym wieku, ma określony ubiór, miejsce w strukturze społecznej, itd.), a te w oczach jednych badanych będą bardzo dobrze pasować do wyobrażonego obrazu „typowego mężczyzny", a w oczach innych – mniej. Innymi słowy, na konstrukt „mężczyzna" będą nakładać się indywidualne cechy danej operacjonalizacji (danego mężczyzny), prowadząc do „zanieczyszczenia" konstruktu. Z tej perspektywy eksperymenty, w których używa się tylko jednej operacjonalizacji jako reprezentacji danego konstruktu, należy uznać za obciążone wadą (ang. *mono-operation bias*) (Shadish, Cook, Campbell, 2002: 75).

Problematyka trafności to jeden ważny punkt widzenia. Innej, ale także istotnej dla omawianych tu zagadnień perspektywy patrzenia, dostarczają koncepcja rzetelności i teoria uniwersalizacji (ang. *generalizability theory*, tłumaczenie terminu za: Brzeziński, 2009) (Cronbach i in., 1972; Shavelson, Webb, 1991; Brennan, 2001). Odpowiadając na pytanie, od czego może zależeć wynik pomiaru wybranej cechy u jednostki, prócz „poziomu tej cechy u tej jednostki" można byłoby wskazać inne czynniki, które dodatkowo mogły zaważyć na takim, a nie innym wyniku. Przykładowo wynik egzaminu sprawdzającego wiedzę danej osoby może być dobrym odzwierciedleniem poziomu jej wiedzy, ale może też zależeć od pytań, na które odpowiadała, osoby oceniającej, okazji określającej warunki, w jakich miał miejsce egzamin, a także interakcji tych czynników. Jeżeli idzie o ocenę czyichś umiejętności, to te dodatkowe czynniki ingerujące w ocenę są społecznie dobrze rozpoznane, podobnie jak praktyki, które mają zaradzić oddziaływaniu tych czynników. Nie bez powodu, przy egzaminach czy konkursach, dąży się do replikowania pomiaru. Jednym z przejawów tego dążenia jest powoływanie kilkuosobowej komisji oceniającej w miejsce jednego egzaminatora. Wpływ egzaminującego na ocenę uwidacznia się w tym, że każdy egzaminator może wystawić nieco inną notę (a duże rozbieżności w notach świadczą o niskiej rzetelności pomiaru). W praktyce społecznej podejmowanym działaniem zaradczym jest w takiej sytuacji wyciągnięcie średniej. Na podobnej zasadzie na egzaminach i konkursach nie poprzestaje się na jednym pytaniu, czy zadaniu, gdyż jest prawdopodobne, że przy kolejnym uczestnik wypadnie lepiej albo gorzej. Tutaj też powiemy, że oceniając odpowiedzi na kolejne pytania (występy w kolejnych konkurencjach) dokonywana jest replika pomiaru umiejętności egzaminowanego i że wyniki tych kolejnych pomiarów nie muszą być takie same. I znowu, aby wpływ pytania zneutralizować, oblicza się sumę albo średnią z not dla kolejnych zadań. Natomiast przeprowadzając

badania metodologiczne na takim materiale (przy założeniu, że byłoby wielu uczestników, co najmniej kilka zadań przypadających na uczestnika, i co najmniej kilku oceniających przypadających na uczestnika i na zadanie), można byłoby ustalić, jaki procent zmienności wyników pochodzi od uczestników (ta zmienność jest pożądana, gdyż odzwierciedla różnice w poziomie umiejętności uczestników), jaki procent pochodzi od oceniających (ta zmienność jest niepożądana, ponieważ wskazuje na zróżnicowany poziom surowości oceniających lub być może zróżnicowane kryteria oceny), jaki procent od zadań (ta zmienność również nie jest pożądana, gdyż świadczy o tym, że zadania różnią się poziomem trudności), a także jaki procent zmienności pochodzi z możliwych interakcji między czynnikami oraz niekontrolowanych błędów pomiarów (ta zmienność też nie jest pożądana). Wyniki takich badań mogłyby być podstawą starań o poprawę rzetelności pomiarów poprzez zmniejszenie rozmiaru zmienności niepożądanych.

Do listy przykładów dołożyć można jeszcze postać kodera występującą w badaniach analizy zawartości przekazów masowych. Koder dokonuje analizy tekstów (na przykład korpusu artykułów prasowych) z wykorzystaniem klucza kodowego. Dawno wypracowana praktyka badawcza zaleca angażowanie kilku koderów (a nie jednego), mając na względzie, że żaden z nich nie jest „maszyną", i że względu na ich światopoglądy, czy inne cechy osobnicze, których nie da się „wyłączyć", każdy z nich może nieco inaczej ocenić ten sam tekst. Z tych też powodów monitoruje się zgodność decyzji koderów w sposobie zakwalifikowania tekstu do poszczególnych kategorii klucza (zob. Lisowska-Magdziarz, 2004).

Podsumowując, w naukach społecznych wiele zmiennych niezależnych stwarza specyficzne problemy pojawiające się przy operacji przejścia od konstruktu teoretycznego do manipulacji. Aplikowanie danej kategorii zmiennej niezależnej w postaci bodźca oznacza co innego, gdy w grę wchodzi taka zmienna, jak płeć czy zachowanie (np. eksperymentatora), a co innego, gdy chodzi o taką zmienną jak warunki fizyczne, zdarzenie X, związek chemiczny, stan fizjologiczny czy emocjonalny. Podobnie manipulacja dystansem przestrzennym, temperaturą, oświetleniem, hałasem, określoną substancją psychoaktywną, programem interwencyjnym X jest bezproblematyczna w tym znaczeniu, że możliwe jest przekazanie bodźca zawierającego wpływ badanej zmiennej niezależnej i tylko tej zmiennej. Jak pokazano na wielu przykładach wyżej, duża grupa zmiennych nie daje takiej możliwości – bodziec nie jest pełną reprezentacją badanego wariantu cechy, a dodatkowo zawiera wpływy pochodzące od innych cech. Problem ten występuje również na gruncie badań nieeksperymentalnych.

Każdy z wymienionych tu przykładowych czynników – nauczyciel, psychoterapeuta, eksperymentator, ankieter, historyjka, reklama, wyraz, komunikat, zadanie, egzaminator, okazja, koder – powinien być uwzględniony w badaniu jako oddzielny czynnik wpływu. Jak pokazano, może on wystąpić w układzie badawczym jako idiosynkratyczny nośnik lub nosiciel bodźca, idiosynkratyczny przypadek kategorii ogólnej, źródło nierzetelności pomiaru, czy – odwołując się do przykładu z podrozdziału 1.1 – okoliczność określająca warunki odkrytej prawidłowości. Najlepiej aby jego oddziaływanie, poprzez odpowiednią aranżację badania (włączenie wielu operacjonalizacji, a nie jednej) oraz uwzględnienie w analizie, zostało odseparowane od działania pozostałych czynników. Jeżeli zaś chodzi o etap analizy, najbardziej pożądanym rozwiązaniem byłoby, aby każdy taki czynnik włączyć do badania jako czynnik losowy (Jackson, Brashers, 1994a; Clark, 1973).

Nawiązując do przedstawionej na początku charakterystyki czynników losowych, a także mając w pamięci konkretne przykłady, warto zwrócić uwagę na fakt, że czynnik losowy jest inną kategorią cechy niż czynnik stały, przez co pełni w badaniu inną rolę. Czynniki stałe występują w funkcji zmiennych eksperymentalnych lub zmiennych kontrolnych. Gdy czynnik stały występuje w roli zmiennej eksperymentalnej (ang. *treatment variable*), to jego wariantami (poziomami) będą poszczególne zabiegi (bodźce) eksperymentalne, które zastosuje badacz wobec badanych. Zabiegi te mogą mieć postać działań, np. metoda nauczania (jak badani reagują na różne metody?) lub być własnościami obiektu, np. wiek nauczyciela (jak badani reagują na nauczycieli w różnym wieku?). Zmienna eksperymentalna odnosi się więc do warunków, którymi manipuluje badacz. Mogą nimi być różne odmiany działania wobec osób badanych albo odmienne własności obiektów. Czynnik stały może też występować w roli zmiennej, którą badacz wprawdzie nie może manipulować w ścisłym sensie, ale chce mieć nad nią kontrolę – zwykle jest to cecha typu własność obiektu i dotyczy osób badanych, np. tożsamość płciowa badanych (czy kobiety reagują tak samo jak mężczyźni?). Na tym tle czynnik losowy jest zmienną nietypową, gdyż jest cechą typu obiekt[4], a jej wariantami (poziomami) są poszczególni przedstawiciele obiektu (danej klasy). Każdy zaś przedstawiciel obiektu jest wiązką niezliczonej liczby własnych cech charakterystycznych. Hasłowo ujmując różnicę między czynnikami stałymi a losowymi, mówi się niekiedy w literaturze, że te pierwsze są tym, czym się manipuluje (w szerokim zna-

[4] Niekiedy grupy obiektów, np. klasy uczniów, szkoły, itd.

czeniu tego słowa), te drugie natomiast są tym, co się losuje lub dobiera do próby (Maxwell, Delaney, Kelly, 2018: 549). Współgra z tym propozycja Colemana, który proponuje, by o każdym czynniku losowym mówić *zmienna generalizacyjna* (ang. *generalization variable*). Zgodnie z intencją autora, termin ten miałby spowodować, że myślenie kategoriami próby i populacji nie ograniczałoby się jedynie do badanych jednostek, ale rozszerzyłoby się na inne obiekty obecne w planie badawczym (takie jak eksperymentator, ankieter, historyjka, reklama itd.) i uświadamiało, że one również objęte są możliwością rozciągnięcia wniosków, ale – tak jak w przypadku badanych jednostek – dopiero po zastosowaniu odpowiednich procedur terenowych oraz analitycznych. Nazwa ta miałaby podkreślać odmienny status zmiennych typu obiekt, a stając się substytutem istniejących nazw ang. *random variable*, *sampling variable*, przenosić akcent z procedury próbkowania na procedurę uogólniania (Coleman, 1972–73).

1.2.3. Plany badawcze bez replikacji i z replikacjami

Wróćmy teraz do przykładu z dwoma metodami nauczania i przeanalizujmy wybory, jakich mógłby dokonać badacz w związku z koniecznością zdecydowania się na jakieś rozwiązanie w odniesieniu do osoby nauczyciela.

Plan badawczy 1a: badacz zatrudnia łącznie dwóch nauczycieli, przy czym nauczyciel n_1 będzie uczył metodą tradycyjną m_1, a nauczyciel n_2 będzie uczył metodą nowatorską m_2. Każdy z uczestników jest uczony tylko jedną metodą i tylko przez jednego nauczyciela (trafił losowo do jednej z dwóch grup). Rysunek 1. przedstawia schemat tabeli danych odpowiadający temu rozwiązaniu. Uczestników ponumerowano według konwencji $s_{numer\ w\ grupie,numer\ grupy}$.

m_1					m_2				
n_1					n_2				
$s_{1,1}$	…	…	…	…	$s_{1,2}$	…	…	…	…
$s_{2,1}$	…	…	…	…	$s_{2,2}$	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	$s_{25,1}$	…	…	…	…	$s_{25,2}$

Rysunek 1. Schemat tabeli danych dla planu 1a

Źródło: opracowanie własne.

Plan badawczy 2a: badacz zatrudnia wielu nauczycieli, powiedzmy, że łącznie dziesięciu[5], przy czym pięciu nauczycieli (n_1–n_5) będzie uczyło metodą m_1, a pięciu kolejnych (n_6–n_{10}) – metodą nowatorską m_2. O ile każdy z dziesięciu nauczycieli potrafiłby uczyć obiema metodami, to metoda nauczania powinna zostać przypisana do nauczyciela losowo (a wcześniej – o ile to możliwe – próba nauczycieli wylosowana z populacji nauczycieli). Natomiast jeśli każdy z nauczycieli potrafiłby uczyć tylko jedną metodą, to preferowanym rozwiązaniem byłoby wylosowanie próby z populacji uczących metodą m_1 oraz wylosowanie kolejnej próby z populacji uczących metodą m_2 (Maxwell, Delaney, Kelly, 2018: 575; Keppel, Wickens, 2004: 551, 559). Każdy z uczestników jest uczony tylko jedną metodą i tylko przez jednego nauczyciela (został przypisany losowo do jednej z dziesięciu grup). Rozwiązanie to ilustruje rysunek 2.

m_1					m_2				
n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}
$s_{1,1}$	$s_{1,2}$	…	…	…	…	…	…	…	$s_{1,10}$
$s_{2,1}$	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
$s_{5,1}$	…	…	…	…	…	…	…	…	$s_{5,10}$

Rysunek 2. Schemat tabeli danych dla planu 2a

Źródło: opracowanie własne.

Plan badawczy 1b: badacz zatrudnia tylko jednego nauczyciela n_1, który będzie uczył zarówno metodą m_1, jak i m_2. Każdy z uczestników uczony jest tylko jedną metodą (trafił losowo do jednej z dwóch grup). Schemat tego rozwiązania pokazuje rysunek 3.

m_1					m_2				
n_1					n_1				
$s_{1,1}$	…	…	…	…	$s_{1,2}$	…	…	…	…
$s_{2,1}$	…	…	…	…	$s_{2,2}$	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	$s_{25,1}$	…	…	…	…	$s_{25,2}$

Rysunek 3. Schemat tabeli danych dla planu 1b

Źródło: opracowanie własne.

[5] Przykład jest uproszczeniem sytuacji badawczej i dobrano do niego niewielkie liczebnie próbki nauczycieli oraz uczniów. W realnym badaniu liczebności próbek powinny być określone ze względu na pożądaną moc testu.

Plan badawczy 2b: badacz zatrudnia wielu nauczycieli, powiedzmy, że łącznie pięciu (n_1–n_5), przy czym każdy z nich będzie uczył zarówno metodą m_1, jak i m_2. W wariancie idealnym nauczyciele stanowią losową próbę nauczycieli. Każdy z uczestników jest uczony tylko jedną metodą i tylko przez jednego nauczyciela (trafił losowo do jednej z dziesięciu grup). Ilustracją tego rozwiązania jest rysunek 4.

m_1					m_2				
n_1	n_2	n_3	n_4	n_5	n_1	n_2	n_3	n_4	n_5
$s_{1,1}$	$s_{1,2}$	…	…	…	…	…	…	…	$s_{1,10}$
$s_{2,1}$	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
$s_{5,1}$	…	…	…	…	…	…	…	…	$s_{5,10}$

Rysunek 4. Schemat tabeli danych dla planu 2b

Źródło: opracowanie własne.

Powyższe rozwiązania nie wyczerpują wszystkich możliwości badawczych. Ograniczają się do schematów międzygrupowych, a więc takich, w których osoba badana jest poddawana tylko jednemu zabiegowi badawczemu (tutaj: jest uczona tylko jedną metodą i przez jednego nauczyciela). Są to schematy kompletnej randomizacji, będące rozszerzeniem planu z grupą kontrolną bez pretestu (Sułek, 1979: 99, 141) (ang. *between-subjects design posttest only*), do którego analizy wykorzystuje się metodę ANOVA (Maxwell, Delaney, Kelly, 2018: 649). Są to jednocześnie schematy popularne w wielu dziedzinach nauk społecznych i choć nie wszystkie są poprawne (do rozważanego tutaj problemu badań), to stanowią cztery schematy prototypowe[6].

[6] Dla tych czterech prototypów proponowano różne nazwy. Jedna z propozycji pochodzi od Jackson, O'Keefe, Brashers (1994). Zgodnie z nią schemat oznaczony jako 1a można określić – w tłumaczeniu adekwatnym, a nie dosłownym – nazwą schemat bez replikacji i bez szablonu (ang. *unreplicated unmatched design*), schemat 1b jako schemat bez replikacji z szablonem (ang. *unreplicated matched design*), schemat 2a jako schemat z replikacjami bez szablonu (ang. *replicated unmatched design*), a schemat 2b jako schemat z replikacjami i z szablonem (ang. *replicated matched design*). Natomiast zgodnie z wcześniejszą, nieco inną propozycją Jackson (1992), schemat 1a określony został jako ang. *unreplicated categorical comparisons*; schemat 1b jako ang. *unreplicated treatment comparisons*; schemat 2a jako ang. *replicated categorical comparisons*; a schemat 2b jako ang. *replicated treatment comparisons*. Każda z tych propozycji wydaje się godna uwagi, niemniej żadna nie funkcjonuje w szerokim obiegu. Na dodatek każda wiąże się z koniecznością wprowadzenia dodatkowych uregulowań definicyjnych. Z tych powodów poprzestałam na rozróżnieniu na schematy z replikacjami i bez

Rozwiązania 2a i 2b można nazwać **schematami z replikacjami**, ponieważ ich cechą charakterystyczną jest to, że na jedną metodę nauczania przypada wielu nauczycieli. Posługując się określeniami z paragrafu 1.2.2, można powiedzieć, że to sytuacja, w której do zaaplikowania bodźca stosowanych jest kilka jego operacjonalizacji czy też kilka ukonkretnionych realizacji bodźca. Można to też ująć jako sytuację, w której eksperyment jest „replikowany" przez każdego kolejnego nauczyciela – eksperyment w zminiaturyzowanej postaci jest powtarzany tyle razy, ilu jest nauczycieli. Rozwiązania 1a i 1b będą natomiast **schematami bez replikacji**, ponieważ na jedną metodę nauczania przypada tylko jeden nauczyciel. Bodziec ma tylko jedną operacjonalizację, stąd sam eksperyment ma tylko jedną realizację[7].

Z wyborem schematu nieuchronnie wiążą się kwestie takie jak: liczebność próbki uczestników, wymogi i ograniczenia organizacyjne czy wymogi związane z właściwą aplikacją bodźca (np. w przypadku nauki języka obcego być może istniałyby zalecenia mówiące, że grupa nie może przekraczać określonej liczby uczestników), itd., niemniej kwestie te tutaj zostały uznane za drugorzędne lub w ogóle pominięte, na rzecz wyeksponowania wadliwości i poprawności poszczególnych rozwiązań, a w przypadku tych poprawnych – ich zalet i ograniczeń.

1.2.3.1. Ocena planu 1a

I tak, rozwiązanie 1a należy uznać za wadliwe i najgorsze z wszystkich czterech przedstawionych. Błąd tkwiący w tym schemacie polega na tym, że oddziaływanie metody m_1 nakłada się na oddziaływanie cech

replikacji, a w dalszej części głównego wywodu zastosowałam nazewnictwo wywodzące się z języka statystyki, które jest bardziej rozpowszechnione. Umożliwia ono wyróżnienie schematu ze strukturą zagnieżdżoną (2a) oraz schematu ze strukturą krzyżową (2b). Ponieważ jednak ta propozycja nie pozwala różnicować schematów bez replikacji, równolegle przyjęłam roboczy podział na schematy 1a, 1b, 2a i 2b.

[7] Termin „replikacje" funkcjonuje w nazewnictwie statystycznym także w innym znaczeniu. Mianowicie nazwy tej używa się również w odniesieniu do pomiarów w grupie eksperymentalnej. Grupy eksperymentalne są wyznaczane przez poziomy czynnika, a jeżeli czynników jest więcej niż jeden, to grupy są wyznaczane przez strukturę poziomów czynników. Struktura ta odzwierciedla krzyżową lub hierarchiczną relację, w jakiej pozostają ze sobą poszczególne czynniki. Jeżeli liczba pomiarów w pojedynczej grupie eksperymentalnej wynosi jeden, to mówimy o schemacie bez replikacji, a jeżeli wynosi co najmniej dwa, to mówimy o schemacie z replikacjami (Kirk 1968, Stanisz 2007, Malarska 2005). Tak rozumianymi schematami bez replikacji będą schematy wewnątrzgrupowe. Przy powyższym znaczeniu schematami z replikacjami będą schematy międzygrupowe poza wyjątkiem, jakim jest schemat międzygrupowy z jedną obserwacją w grupie (ang. *completely randomized factorial design with n = 1*, Kirk 1968: 227).

nauczyciela n_1, a oddziaływanie metody m_2 nakłada się na oddziaływanie cech nauczyciela n_2. Nakładanie się oddziaływań czynników, których nie można rozdzielić prowadzi do analogicznego problemu, który stwarzała wcześniej omówiona tzw. selekcja, zagrażająca wewnętrznej trafności eksperymentu. Tutaj ten problem oddaje wątpliwość: jak wyjaśnić różnicę (lub jej brak) między wynikami w teście w grupie m_1 i m_2? Czy za obserwowanym efektem stoi metoda nauczania, nauczyciel czy może jeden i drugi czynnik? (Jackson, Brashers, 1994a: V, 3; Keppel, Wickens, 2004: 551; Shavelson, Webb, 1991: 47–48). Być może dodatkowo pomocny będzie dla Czytelników rzut oka na rysunek 1. i próba uzmysłowienia sobie, co by się stało, gdyby tworząc elektroniczny zbiór danych dla tego schematu, wprowadzić oddzielne zmienne dla czynnika „metoda" i dla czynnika „nauczyciel" – okazałoby się, że każda z tych zmiennych powiela informacje drugiej.

Omawiany błąd rozwiązania 1a jest różnie nazywany. Opierając się na szeroko rozpowszechnionej typologii trafności Cooka i Campbella (1979), część autorów odnosi go do trafności wewnętrznej (na przykład: Rumenik, Capasso, Hendrick, 1977; Jackson, Jacobs, 1983; Brashers, Jackson, 1999; Jackson, O'Keefe, Jacobs, 1988), część z kolei do trafności teoretycznej (Wells, Windschitl, 1999). Jeżeli przyjmiemy, że wszędzie tam, gdzie nakładanie się oddziaływań następuje w wyniku czynności operacyjnych (ang. *construct confounding*, Shadish, Cook, Campbell, 2002: 75) mamy do czynienia z problemem związanym z trafnością teoretyczną, to przypadek rozwiązania 1a powinien być podciągnięty pod tę kategorię. Problem polega jednak na tym, że taka kwalifikacja uniemożliwiałaby zniuansowanie wszystkich defektów rozwiązania 1a, a także odróżnienie defektów rozwiązań 1a i 1b. W tej sytuacji „zakłócanie konstruktu", czyli jeden z możliwych czynników upośledzających trafność teoretyczną, wydaje się kategorią zbyt szeroką i przez to mało użyteczną, szczególnie do celów analizy metodologicznej. Kategoria ta mieści w sobie problemy typu „nakładanie się oddziaływań" (wiązane definicyjnie z trafnością wewnętrzną) oraz problemy dotyczące uogólniania zaobserwowanej prawidłowości (wiązane definicyjnie z trafnością zewnętrzną). Zatarcie tych różnic tylko ze względu na element wspólny: „…w wyniku czynności operacyjnych" jest niekorzystne. Optowałabym raczej za ujęciem, zgodnie z którym problem trafności teoretycznej leży u podłoża sprawy i w zależności od tego, jaki plan badawczy zostanie przyjęty, może przerodzić się on w problem trafności wewnętrznej i/lub zewnętrznej. Ostatecznie omówiony dotąd błąd rozwiązania 1a ujmuję jako zakłócający trafność wewnętrzną.

Można dodać, że w literaturze proponowano jeszcze inne nazwy dla tego błędu, takie jak ang. *category-confound* (Kay, Richter, 1977), czy ang. *case-category confounding* (Jackson, 1992: 31). Wprawdzie propozycje te sformułowano poza klasyczną typologią trafności Cooka i Campbella, ale ich wartością jest to, że dobrze obrazują defekt.

Rozwiązanie 1a – mimo iż obciążone wadą – zamieszczono w tekście, gdyż przy małolicznej próbie badawczej może wydawać się uzasadnione i na swój sposób oczywiste. Co więcej, mimo swej wadliwości, jest ono dość powszechnie stosowane, choć równie często krytykowane. Przykładowo listę studiów obarczonych tą wadą w dziedzinie komunikacji społecznej Czytelnik znajdzie u Jackson i Jacobs (1983).

Najwięcej uwagi dla ułomności tego rozwiązania poświęcili bodaj Wells i Winschtil (1999). Wskazywali, że nieraz defekty tego rozwiązania są rozpoznawane jako oczywiste, ale w sytuacjach mniej ewidentnych bądź ugruntowanych przez złe praktyki badawcze stają się mniej dostrzegane. Gdyby więc iść tropem autorów i przerysować problem, można byłoby podać przykład eksperymentu, w którym badacza interesuje, które osoby są postrzegane jako bardziej inteligentne – o imieniu jedno czy wielosylabowym. Dla osiągnięcia tego celu osoby badane zostałyby poproszone o ocenę tego samego eseju – przy czym osoby w jednej grupie dowiedziałyby się, że jego autorem był Lech, a w drugiej Aleksander (adaptacja przykładu – KGR). Przykład ten – być może dodatkowo rażący poznawczą miałkością problemu – dobrze oddaje istotę omawianego tu defektu. Wprawdzie oba imiona są męskie (różnicy w przypisywanym stopniu inteligencji nie można tłumaczyć płcią), ale nie jest tak, że różnią się jedynie liczbą sylab – różnią się częstością występowania, frekwencją nadawania tych imion dzieciom obecnie urodzonym, częstością obecności w literaturze i historii, skojarzeniami z osobami publicznymi, w tym politykami noszącymi te imiona, itd. Pokaźna liczba łatwo uchwytnych różnic między tymi imionami w połączeniu z informacją, co stanowiło cechę odróżniającą imiona w zamyśle badacza (dla większości odbiorców jest to pewnie cecha peryferyjna i być może nawet poproszeni o samodzielne wskazanie różnic między tymi imionami nie odgadliby jej) pokazują, że do rozwiązania typu 1a należy podchodzić z dużym sceptycyzmem. Pojedyncza operacjonalizacja danej kategorii nie może być jej dobrym reprezentantem – jest przedstawicielem wielu kategorii i sama żadnej pojedynczej kategorii nie reprezentuje wystarczająco dobrze.

Wells i Winschtil (1999) wskazywali dalej, że wadliwość omawianego rozwiązania jest już zdecydowanie mniej oczywista, gdy badany jest efekt płci czy rasy. Tutaj większa część praktyków-badaczy jest skłonna uznać, że wystarczy, by płeć żeńską reprezentowała jedna

kobieta, a płeć męską jeden mężczyzna (analogicznie sprawa przedstawia się w przypadku koloru skóry). Po taki schemat badawczy sięgnęła 1/3 badaczy interesujących się wpływem płci eksperymentatora (Rumenik, Capasso, Hendrick, 1977). Występują też przykłady użycia omawianego schematu na gruncie badań sondażowych – efekt płci ankietera ustalali w ten sposób chociażby Landis, Sullivan, Sheley (1973), czy Fuchs (2009), choć nie bez autokrytyki dla zastosowanego podejścia. Być może jest tak, że najwyraźniej rzadsze wykorzystywanie tego schematu na polu badań sondażowych w porównaniu z polem eksperymentów psychologicznych jest po prostu pochodną specyfiki każdego z tych typów badań (eksperymenty psychologiczne prowadzi się zwykle na małych próbach i do ich „obsługi" może wystarczyć co najwyżej kilkuosobowy zespół zbierający dane, natomiast w przypadku badań sondażowych, w tym także eksperymentów sondażowych, zwykle korzysta się z dużych prób i większej liczby ankieterów).

Dla obrońców schematu 1a argumentem niepozbawionym racji jest to, że kategoryzowanie w oparciu o płeć, podobnie jak w oparciu o kolor skóry, jest zdecydowanie łatwiejsze niż grupowanie w oparciu o jakąkolwiek inną cechę, w tym liczbę sylab w imieniu (Wells, Winschtil, 1999). Dlatego słysząc imiona takie jak Roch i Maria, badani najpewniej wiązaliby je w pierwszym rzędzie z różnicą płci, a z różną liczbą sylab być może wcale. Nawet w takiej sytuacji nie można byłoby wyciągnąć wniosku, że różnica płci wyczerpuje w oczach badanych różnice między tymi imionami. Podobnie, widząc przedstawiciela płci męskiej i przedstawicielkę płci żeńskiej, badani nie musieliby upatrywać w płci jedynej różnicy między tymi osobami, ale dodatkowo widzieć – na przykład – różnice klasowe, czy jeszcze inne. Decydując się na rozwiązanie, w którym dobieramy tylko dwie operacjonalizacje – jedną, by reprezentowała pierwszą kategorię cechy X i drugą, by reprezentowała drugą kategorię cechy X – nie możemy liczyć na to, że operacjonalizacje te będą się różniły tylko ze względu na cechę X. Wyobrażenie, że spełnione zostały warunki kanonu jedynej różnicy jest w tej sytuacji złudzeniem.

Jedynym poprawnym postępowaniem w sytuacji, w której doszło do posłużenia się schematem 1a jest zmiana interpretacji, która polegałaby na tym, by o pojedynczych operacjonalizacjach mówić nie jako reprezentantach ogólnych kategorii, ale jako idiosynkratycznych jednostkach składających się z wielu własnych cech charakterystycznych, a także oznaczać je w sposób, który by to oddawał – w przypadku osób, na przykład ich własnym imieniem i nazwiskiem. Dlatego zamiast pisać w raporcie wniosek przykładowo: „badani byli bardziej reaktywni,

gdy złą wiadomość przekazywał im mężczyzna niż gdy przekazywała ją kobieta", należałoby napisać „badani byli bardziej reaktywni, gdy złą wiadomość przekazywał im współpracownik badacza Piotr G. niż gdy przekazywała ją współpracowniczka badacza Agnieszka T.". W raportach jednak trudno znaleźć zdanie drugiego typu. Dzieje się tak najprawdopodobniej dlatego, że zdanie to przekazuje sugestię, iż efekt płci jest niepewny i niestabilny – inne pary współpracowników być może spowodowałyby, że wyniki byłyby inne. Tym samym zdanie to odsłania kolejną słabość schematu 1a – mianowicie to, że odkryta „prawidłowość" dotyczy raptem pary osób, nie wiadomo czy innych również, a zatem jest pozbawiona większej wartości poznawczej. Innymi słowy, zdanie to demaskuje bardzo niską trafność zewnętrzną ustalenia badawczego. Przywołajmy definicję – wysoka **trafność zewnętrzna** oznacza, że prawidłowość utrzymuje się: 1) przy różnicowaniu osób badanych, warunków badawczych, wariantów zabiegu i sposobów pomiaru, które były wykorzystane w badaniu, a także utrzymuje się dla 2) osób, warunków, wariantów zabiegu i sposobów pomiaru, które nie były wykorzystane w badaniu (Shadish, Cook, Campbell, 2002: 83). W omawianym tu rozwiązaniu 1a uwaga dotycząca trafności zewnętrznej odnosi się konkretnie do braku wariantowości zabiegu badawczego. Dodajmy jeszcze i tę oczywistą rzecz, że gdyby nawet próbować podnieść poziom trafności zewnętrznej poprzez zwiększenie próby uczestników, to i tak zabieg ten na niewiele się zda – pozwoli wprawdzie ocenić, czy zaobserwowany wzór reakcji utrzymuje się wobec większej liczebnie (i być może bardziej zróżnicowanej) zbiorowości osób, niemniej będzie to cały czas wzór reakcji na Piotra G. i Agnieszkę T.

Dlaczego jest tak, że jako badacze nie zadowalamy się efektem stwierdzonym wobec jednego uczestnika ($n = 1$), uznajemy, że efekt ten może być jednostkowy i nie reprezentuje właściwie populacji N, a jednocześnie jesteśmy skłonni uznać, że jeden przedstawiciel kategorii ($m = 1$) może sensownie reprezentować całą kategorię M (wszystkich jej przedstawicieli)? Dlaczego badając efekt płci czy koloru skóry, dobieramy tylko po jednym przypadku z kategorii? Według hipotezy Wellsa i Winschtila (1999) dzieje się tak dlatego, że badacze – podobnie jak inni ludzie – ulegają błędom poznawczym. Po pierwsze chodzi o uleganie „prawu małych liczb". Zgodnie z koncepcją Tversky'ego i Kahnemana (1971) ten rodzaj skrzywienia poznawczego polega na uznaniu, że charakterystyka dużych prób losowych, w szczególności ich duży poziom reprezentatywności, stosuje się także do małych prób. Po drugie, jak argumentują Wells i Winschtil (1999), chodzi o uleganie heurystyce reprezentatywności. Potrzebę zbudowania próby z większej liczby

przypadków jesteśmy – tak jak inni ludzie – skłonni postulować wtedy, gdy dostrzegamy zróżnicowanie w obrębie przypadków, ale jeśli tylko występuje pewnego rodzaju stereotyp kategorii, to przysłania nam on to zróżnicowanie i zaczyna wyznaczać (ujednolicony) obraz tego, co „reprezentatywne". Tymczasem stereotypowy obraz niekoniecznie musi odpowiadać tendencji centralnej (tutaj dominancie).

Gdyby zastosowanie jednej operacjonalizacji (dla jednej kategorii) miało zostać właściwie uzasadnione, wymagałoby to wcześniejszego przebadania większej liczby operacjonalizacji i udokumentowania, że badani reagują na każdą z nich w ten sam sposób (wtedy tylko jedna z nich mogłaby być wybrana, nieważne która). Powodzenie tego przedsięwzięcia jest niepewne i nie ma dlań „planu b".

Pozornie obiecująca strategia mogłaby też polegać na próbie kontrolowania maksymalnie dużej liczby cech dwóch operacjonalizacji tak, aby różniły się tylko pod względem jednej cechy. Zdaniem Colemana (1979: 171) nigdy nie można mieć pewności, czy wszystkie istotne cechy udało się utrzymać na jednym poziomie (także: Jackson, O'Keefe, Jacobs, 1988). Z kolei przykład z imionami pokazuje, że mogą występować materie niepodatne na takie operacje.

Zamiast wkładać trud w niepewne przedsięwzięcia, zdecydowanie lepiej jest spożytkować siły inaczej – włączyć próbę operacjonalizacji do badania zasadniczego. Badanie to nie tylko pozwoliłoby zdobyć informację o tym, czy i w jakim stopniu operacjonalizacje reprezentujące daną kategorię różnią się między sobą, ale umożliwiłoby również – przy zastosowaniu procedur wnioskowania statystycznego – generalizację wniosków z próby przypadków na populację przypadków (na przypadki tworzące kategorię).

Badacze korzystający ze schematu 1a zwykle zostawiają czytelników z niedopowiedzeniem odnośnie zakresu wniosków. Co najwyżej robią krótkie zastrzeżenia końcowe – piszą, że każda z badanych przez nich kategorii jest reprezentowana przez jeden przypadek, postulują potrzebę dalszych badań i zostawiają w zawieszeniu kwestię, czy ustalenie badawcze stosuje się do wszystkich przypadków kategorii. Tymczasem już Coleman postawił sprawę jasno, pisząc z przekąsem, cyt.: „*Jeżeli eksperymentator sugeruje, że jego odkrycie można uogólnić na konkretną populację, nie jest nierozsądne wymaganie od niego statystycznego dowodu, że to prawda*" (1972–73: 227). Sugestie, że na podstawie $m = 1$ można wnioskować o M należy zdecydowanie uznać za metodologicznie nieuprawnione, a z punktu widzenia rozwoju wiedzy – niekorzystne. Studia badawcze prowadzone na ten sam temat i w oparciu o schemat 1a z reguły przynoszą niespójne rezultaty – i to jest jeszcze jeden dowód na

to, że poszczególne operacjonalizacje różnią się efektem, a zatem tylko włączenie większej liczby operacjonalizacji do fazy terenowej i analitycznej jest sensownym rozwiązaniem.

1.2.3.2. Ocena planu 1b

Wróćmy do przykładu badania z nauczycielami i rozpatrzmy rozwiązanie 1b. Plan ten przewiduje zatrudnienie jednego nauczyciela znającego dwie metody nauczania (metodą m_1 będzie uczył jedną grupę, a metodą m_2 – drugą grupę badanych). Oczywiście nie zawsze możliwie jest zastosowanie takiego schematu – analizowany tu przykład akurat tego nie uwyraźnia, bowiem opanowanie kilku metod nauczania przez nauczyciela, jakkolwiek może stanowić pewną trudność organizacyjną badania, jest wykonalne. Inaczej sytuacja wyglądałaby, gdyby czynnikiem stałym nie miałaby być „metoda nauczania", ale jakaś niezmienna charakterystyka osobnicza (tu nauczycieli), czy też cecha niepodatna na przełączanie się z jednego wariantu w drugi, np. „cecha osobowości", czy „tożsamość płciowa" (u osób cispłciowych).

Gdyby porównać schemat 1b z poprzednio omówionym schematem 1a, to wynik tego porównania – co do zasady – wypadłby na korzyść 1b. Schemat ten – potencjalnie przynajmniej – zabezpiecza przed zakłócaniem oddziaływania czynnika „metoda" oddziaływaniem czynnika „nauczyciel", jako że ten ostatni utrzymywany jest na tym samym poziomie. Mówiąc językiem badań eksperymentalnych, rozwiązanie 1b nie zagraża trafności wewnętrznej eksperymentu, ponieważ nie jest tak, że na m_1 nakłada się n_1, a na m_2 nakłada się n_2. Tutaj, zarówno na poziomie m_1, jak i na poziomie m_2 oddziałuje ten sam nauczyciel n_1. Widzimy tu drugi – obok randomizacji – sposób radzenia sobie z czynnikami zakłócającymi. Sposób ten, skądinąd najprostszy, polega na utrzymywaniu czynnika zakłócającego na stałym poziomie (Keppel, Wickens, 2004: 6).

Powodzenie tego schematu, jego stopień trafności wewnętrznej, ostatecznie zależy od tego, czy zabiegi badawcze rzeczywiście różnią się tylko pod względem jednej zmiennej (zob. Brashers, Jackson, 1999: 462–463). W omawianym przykładzie chodziłoby o to, aby grupy m_1 i m_2 rzeczywiście różniły się tylko metodami nauczania. Drobiazgowe przyjrzenie się temu schematowi nakazywałoby zadać pytanie, czy nauczyciel n_1 – poza różnicą w nauczanej metodzie – demonstruje dokładnie te same zachowania w grupach m_1 i m_2. Innymi słowy, na ile mocna jest podstawa, że czynnik „nauczyciel" jest rzeczywiście utrzymywany na tym samym poziomie oraz czy nie mamy przypadkiem do czynienia z kontrolą pozorną i zbyt pochopnie przyjętym założeniem, że skoro

mamy jedną osobę, to w kolejnych warunkach badawczych będzie przejawiała te same zachowania poza tymi, które badacz chce by się zmieniały. To drobiazgowe i krytyczne przyjrzenie się schematowi 1b nakazuje też zadać pytanie, czy czynniki inne niż „nauczyciel" są kontrolowane (np. czy zajęcia dla grupy m_1 i m_2 rozpoczynają się w tych samych godzinach, czy trwają równie długo, itd.).

Być może są takie rodzaje obiektów, które – w połączeniu ze sprzyjającym pytaniem problemowym – pozwalają na kontrolowaną zmianę w zakresie tylko jednej charakterystyki, a badania z ich udziałem można zasadnie nazwać schematem 1b, ale z pewnością nie dotyczy to wszystkich typów obiektów. Jak obszernie argumentują Jackson, O'Keffe, Jacobs i Brashers (1989), schematu 1b praktycznie nie da się zastosować wobec obiektów typu komunikat. Zmiana komunikatu pod kątem badanej cechy niemal zawsze pociąga za sobą zmianę co najmniej jednej cechy niebadanej (i niekontrolowanej). A więc nawet jeśli badacze przystępują do badania z myślą zrealizowania go w schemacie 1b i kończą je z takim przeświadczeniem, to wyobrażenie to jest iluzoryczne. Wykorzystanie potencjału wysokiej trafności wewnętrznej schematu 1b często nie jest możliwe – pozorne, ułomne schematy 1b mają tę samą niską trafność wewnętrzną, co schematy 1a.

Przyjrzyjmy się drugiemu kryterium dobroci eksperymentu – trafności zewnętrznej, czyli tej charakterystyce, która mówi o ogólności jego ustaleń. Tutaj mamy do czynienia z sytuacją, w której oddziaływania nauczyciela n_1 nie można oddzielić od zaobserwowanego efektu (różnica między m_1 i m_2 jest efektem dla n_1), a zatem stajemy przed pytaniem: czy na podstawie badania, które przeprowadzono z konkretnym, jednym nauczycielem, można wypowiadać się o ewentualnym prymacie jednej metody nad drugą? Czy są zasadne podstawy by uznać, że tę samą konkluzję uzyskanoby angażując innego nauczyciela? Jak już wcześniej argumentowano, nauczyciel nie jest nośnikiem jedynie metody, dlatego założenie, że nauczyciel „nie ma znaczenia", w tym sensie, że może być zastępowany przez innego, wydaje się być wątpliwe. W konsekwencji uczciwiej i bezpieczniej jest traktować osiągnięte wyniki dotyczące porównywanych metod jako ustalenia odnoszące się do jednego, konkretnego nauczyciela, a roszczenie generalizacji wniosków wstrzymać do czasu przebadania większej grupy nauczycieli (Keppel, Wickens, 2004: 548).

Ostatecznie pod względem trafności zewnętrznej rozwiązanie 1b wypada słabo, ponieważ dzieli wszystkie słabości schematu 1a, dlatego całą wcześniejszą argumentację na ten temat można odnieść także do obecnie omawianego rozwiązania.

W odniesieniu do rozwiązań 1a i 1b, warto zaznaczyć jeszcze jedną kwestię: nie każde badanie jednego przypadku jest z gruntu niepoprawne. Istnieją badania, których nie prowadzi się z zamiarem, by analizowany przypadek miał reprezentować ogólną kategorię, i te wyłączyć należy z krytyki (Clark, 1973; Brashers, Jackson, 1999). Przykładem takiego studium byłoby badanie wpływu kanału komunikacji na percepcję postaci Ronalda Reagana i Waltera Mondale'a, walczących w kampanii wyborczej o urząd prezydenta w 1984 roku. Gdyby badacz uprawiał podejście idiograficzne, odnosił się do danej kampanii jako swoistego zdarzenia w historii, postacie Reagana i Mondale'a byłyby dla niego ważne same w sobie, to niecelowe byłoby włączanie do badania komunikatów pochodzących od innych osób startujących w tamtych wyborach. W takiej sytuacji uzasadnione byłoby, aby zmienna „kandydat" przyjęła tylko dwa poziomy i była uznana za czynnik stały (Jackson, O'Keefe, Brashers, 1994).

1.2.3.3. Ocena planu 2b

Pożądanym rozszerzeniem planu 1b będzie rozwiązanie 2b, przewidujące zaangażowanie co najmniej kilku nauczycieli, z których każdy będzie uczył jedną i drugą metodą. Rozwiązanie 2b należy zresztą uznać za najkorzystniejsze spośród wszystkich analizowanych. Zatrudnienie większej liczby nauczycieli daje badaczowi poczucie, że uniezależnia wyniki od cech jednego nauczyciela (w odróżnieniu od schematu 1b), a gdyby dodatkowo nauczyciele stanowili losową próbkę, badacz mógłby się czuć uprawniony do uogólnienia swoich wniosków w zakresie metod nauczania nie tylko na nieprzebadanych uczniów (populację uczniów), ale także na nieprzebadanych nauczycieli (populację nauczycieli). Należy jednak wyraźnie podkreślić, że samo wprowadzenie próbki nauczycieli do etapu terenowego nie wystarczy – dopiero wzbogacenie modelu analitycznego o losowy czynnik „nauczyciel" czyni całość postępowania badawczego w pełni poprawnym, a także pozwala na uogólnienie wyników poza przebadanych nauczycieli. To ważny moment wywodu. Mogłoby się bowiem wydawać, że już tylko przebadanie większej liczby nauczycieli, samo dysponowanie „bardziej zróżnicowanymi" danymi, w wystarczający sposób ulepsza rozwiązanie z jednym nauczycielem. Liberalne podejście, które reprezentują Shadish, Cook i Campbell (2002: 75–76) zaleca wykorzystanie wielu replikacji, czyli zaangażowanie wielu nauczycieli zamiast jednego oraz włączenie czynnika „nauczyciel" do analiz (choć już nie precyzuje w jakim charakterze – czynnika stałego czy losowego). Natomiast przy małolicznej próbie uczestników autorzy ci dopuszczają wariant, w którym na etapie analitycznym ignoruje się fakt,

że dane pochodzą od kilku nauczycieli (przeprowadza się analizę wariancji z jednym czynnikiem „metoda"). Jednak z punktu widzenia poprawności metodologicznej, to ostatnie rozwiązanie nie jest właściwe – jeżeli dane zebrane są od wielu nauczycieli (innymi słowy, z wykorzystaniem replikacji) to powstaje ryzyko, że pomiary przestaną być niezależne. Zagrożenie to ujmowane jest jako brak niezależności z powodu grup[8] (ang. *nonindependence due to groups*), prowadzące do przeszacowania albo niedoszacowania prawdopodobieństwa popełnienia błędu I rodzaju (Kenny, Judd, 1986; Judd, McClelland, Culhane, 1995). Odnieśmy ten problem do analizowanego przykładu. W zależności od tego, gdzie występują korelacje – czy podobne do siebie są pomiary wewnątrz grup wyróżnionych ze względu na metody, czy też podobne do siebie są pomiary wewnątrz grup wyróżnionych ze względu na nauczycieli – można oczekiwać odpowiednio przeszacowania efektu metody (test zbyt liberalny) albo jego niedoszacowania (test zbyt konserwatywny) (zob. Brashers, Jackson, 1999: 464).

Zakładając teraz, że replikacje (tutaj: nauczyciele) są włączone do analiz jako czynnik, przyjrzyjmy się teraz bliżej powstałemu modelowi. W schemacie 2b sposób powiązania poziomów czynnika „nauczyciel" i czynnika „metoda" jest taki, że wszystkie poziomy czynnika „nauczyciel" występują zarówno na poziomie m_1, jak i na m_2, a więc występują na wszystkich poziomach czynnika „metoda". Zachodzi też relacja odwrotna. Każdy nauczyciel uczy wszystkimi (obiema) metodami i każda metoda jest stosowana przez wszystkich (pięciu) nauczycieli. Poziomy obu czynników mogą zostać „skrzyżowane" (w efekcie czego powstanie 10 grup), stąd strukturę, jaką tworzą, określa się w języku statystycznym mianem **krzyżowej**[9] (zob. rysunek 4.). Przy włączeniu do analiz czynnika „metoda" oraz czynnika „nauczyciel", struktura ta umożliwia odseparowanie zmienności, które pochodzą od: 1) metody; 2) nauczycieli; 3) interakcji tych czynników (czy efekt metody jest taki sam dla każdego nauczyciela lub czy efekt nauczyciela jest taki sam dla m_1 i m_2?); 4) badanych osób, przy czym tej zmienności nie da się oddzielić od zmienności wynikającej z działania innych, niekontrolowanych czynników[10]. Schemat 2b, z uwagi na możliwość wejścia ze sobą w interakcję czynnika odnoszącego do zabiegu badawczego (tutaj: „metoda") i czynnika odnoszącego się replikacji (tutaj: „nauczyciele"), okre-

[8] Tutaj w wariancie dla schematów ze strukturą krzyżową.

[9] Schematy oparte na strukturze krzyżowej czynników nazywane są również schematami czynnikowymi (ang. *factorial designs*) (Keppel, Wickens, 2004: 11).

[10] Sposób wyłaniania źródeł zmienności omawiam dokładniej w podrozdziale 2.2.

ślony jest w literaturze schematem ***zabieg przez replikacje*** (ang. *treatment by replication design*), w skrócie $\boldsymbol{T} \times \boldsymbol{r}$.

Do rozpatrzenia pozostaje kwestia statusu czynnika „replikacje" – stały czy losowy? Zwłaszcza w sytuacji, w której próba nauczycieli byłaby losowa, nie byłoby żadnego uzasadnienia, aby czynnik „nauczyciele" traktować jako stały. Zamiary badacza (uogólnienie wniosków na nieprzebadanych nauczycieli) i charakterystyka czynnika stałego (niemożność uogólniania wyników na nieprzebadanych nauczycieli) stałyby ze sobą w sprzeczności. Choć ta sprzeczność wydaje się oczywista, swego czasu przez wielu badaczy nie była w ogóle dostrzegana. Jak pisał jeszcze Clark (1973), początkujący badacze nie mają wiedzy na temat tego, jak należy przeprowadzać uogólnianie wniosków na dwie populacje jednocześnie – działają tak, jak gdyby badani uczestnicy mogli być jedynym efektem losowym w badaniu. *Language-as-fixed-effect fallacy* (nazwa błędu, który określił Clark, a która niesłychanie upowszechniła się w piśmiennictwie naukowym i pączkowała jako *experimenter-as-fixed-effect fallacy*, *therapist as-fixed-effect fallacy*, itd.), a więc błąd polegający na tym, że czynnik, który jest losowy traktuje się jako stały, miał brać się z niewiedzy, czym jest czynnik stały (a czym losowy) i jakie są konsekwencje takiej błędnej kwalifikacji. Nieświadomi tego badacze, uwzględniający jedynie respondentów jako czynnik losowy, błędnie uznawali, że ustalenie badawcze stosuje się do nowych prób respondentów *oraz* do nowych prób replikacji (w analizowanym przykładzie: nowych prób nauczycieli). Tymczasem powyższa kwalifikacja pozwala na uogólnienie węższe – uprawnia badacza do stwierdzenia, że ustalenie badawcze zaobserwowane na jednej, konkretnej próbie replikacji (nauczycieli) stosuje się jedynie do nowych prób respondentów (utrzymałoby się, gdy próbę respondentów miały tworzyć inne osoby).

Bardziej problematyczny przypadek przedstawia sytuacja, w której nauczyciele nie stanowiliby losowej próbki – zostaliby wybrani w sposób celowy, czy też z uwagi na ich łatwą dostępność, itp. Dotykamy tu ówcześnie (i być może współcześnie również) kontrowersyjnej kwestii, czy dozwolone jest stosowanie wnioskowania statystycznego w sytuacji, gdy próba nie została pobrana losowo. Zdaniem Clarka, a także kontynuatorów jego myśli, należących w latach 70. i 80. do awangardy metodologicznej, dziś już raczej przedstawicieli „głównego nurtu", czynnik „replikacje" należy uznać za losowy także wówczas, gdy próbę tworzą replikacje niepobrane losowo. Podstawą kwalifikacji czynnika jako stałego albo losowego nie powinno już być tradycyjne kryterium, jakim jest sposób *wybrania* poziomów czynnika. Głównym kryterium powinien

być sposób i cel *wykorzystania* tych poziomów (Clark, 1973: 348; Jackson, Jacobs, 1999: 170).

Przypomnijmy, że jednym z celów wykorzystania wielu replikacji jest podniesienie trafności wewnętrznej. Pod tym względem nie tylko schemat 2b, ale także jeszcze nieomówiony schemat 2a, przewyższają rozwiązanie 1b, a w szczególności rozwiązanie 1a. Drugim celem – i na niego chcę teraz zwrócić uwagę – jest podniesienie trafności zewnętrznej, a konkretnie zwiększenie zakresu stosowalności wniosku (dotyczącego efektu zabiegu badawczego) w odniesieniu nie tyle do osób badanych, bo wnioskowanie statystyczne w tym zakresie gwarantuje każdy typ ANOVA, ile w stosunku do nauczycieli. Gdyby replikacje miały być uznane za czynnik stały, znaczyłoby to, że ustalenie badawcze należy ograniczyć do przebadanych nauczycieli. Porównując schemat 1b (z jednym nauczycielem) ze schematem 2b, (dla którego przykład przewiduje pięciu nauczycieli), trzeba byłoby przyznać, że przyrost trafności zewnętrznej byłby znikomy, a sam poziom tej trafności ciągle bardzo niski. Z naukowego punktu widzenia ustalenia o takim poziomie ogólności są mało wartościowe. Badacze są świadomi tych ograniczeń, dlatego słusznie zainteresowani są osiągnięciem większego poziomu generalizacji. O ile więc badacz przejawia intencję szerszej generalizacji, użyte przez niego replikacje nie wyczerpują populacji i mogłyby być zastąpione przez inną próbkę, to – niezależnie od tego, czy zostały wybrane losowo, czy nie – badacz powinien potraktować te replikacje jako poziomy czynnika losowego (Clark, 1973: 348–349). Badacze uznają próbę badanych osób jako źródło losowej zmienności nawet wtedy, gdy próba uczestników nie jest losowa, co w badaniach eksperymentalnych zdarza się bardzo często. Na podobnej zasadzie należy potraktować próbę replikacji (Jackson, Brashers, 1994b). Kwalifikacja replikacji jako czynnika losowego umożliwi przeprowadzenie dla replikacji „oddzielnego" wnioskowania statystycznego – analogicznego do wnioskowania, jakie przeprowadza się wobec badanych jednostek z myślą o tym, by wnioski objęły populację jednostek.

Dlaczego błędem byłoby w powyższej sytuacji potraktowanie replikacji jako poziomów czynnika stałego? Odpowiedź jest następująca – bo testując efekt metody, zaszłaby niekompatybilność między hipotezą zerową, która w rzeczywistości jest testowana a hipotezą, którą badacz chciałby testować. Jeżeli potraktujemy replikacje jako czynnik stały, to hipoteza zerowa mówi, że efekt metody dla przebadanych nauczycieli w populacji uczestników jest równy zero. Tymczasem hipoteza, którą badacz jest *implicite* zainteresowany mówi, że efekt metody w populacji nauczycieli oraz w populacji uczestników jest równy zero (Clark, 1973).

Jednocześnie badacz, który zainteresowany jest ostatnią hipotezą, ale traktuje replikacje jako czynnik stały, to formułuje wnioski na podstawie testu, w którym ryzyko popełnienia błędu I rodzaju może być większe niż *alfa*, a więc w warunkach nieuczciwie sprzyjających odrzuceniu hipotezy zerowej. Innymi słowy, przy uznaniu replikacji za czynnik stały może być łatwiej wykazać efekt metody niż wtedy, gdy czynnik ten uznamy za losowy.

Jak unaoczniły to symulacje Monte Carlo, inflacji błędu I rodzaju będzie sprzyjać sytuacja, w której efekt metody będzie zmieniał się w zależności od nauczyciela. Zmienność efektu w zależności od replikacji to w języku statystycznym wystąpienie interakcji między zabiegiem a replikacją. W przypadku takiej interakcji, im większa będzie liczba badanych przypadająca na replikację (tu: nauczyciela), tym większe będzie ryzyko błędu. Utrzymaniu wielkości błędu I rodzaju na poziomie *alfa* = 0.05 będzie sprzyjać duża liczba replikacji, pod warunkiem, że efekt interakcji nie będzie zbyt duży (Jackson, Brashers, 1994b: 367).

Patrząc na to z innej perspektywy, istotny statystycznie efekt zmiennej eksperymentalnej T w przypadku, gdy replikacje traktowane są jako czynnik stały, może pojawić się na skutek trzech różnych sytuacji: 1) w populacji efekt dla zmiennej eksperymentalnej T jest różny od zera i efekt interakcji $T \times r$ jest równy zero, 2) efekt dla zmiennej eksperymentalnej T jest równy zero, ale efekt interakcji $T \times r$ jest różny od zera, 3) obydwa efekty są różne od zera (Clark, 1973; Brashers, Jackson, 1999). Mówiąc swobodnie – skoro traktujemy replikacje jako czynnik stały, a zatem badamy efekt metody w odniesieniu do bardzo konkretnych nauczycieli (a nie do wszystkich nauczycieli), to nie możemy być pewni, czego zaobserwowany efekt jest świadectwem – czy odzwierciedla rzeczywiste różnice między metodami, czy też odzwierciedla indywidualne oddziaływania nauczycieli, z których każdy jest inny, czy też może odzwierciedla obydwa powyższe oddziaływania występujące jednocześnie.

Jak pokazało wiele empirycznych badań – głównie w dziedzinie komunikacji społecznej, gdzie schemat 2b ma szerokie zastosowanie – wystąpienie interakcji, a więc sytuacja, w której poszczególne replikacje (komunikaty) różnią się badanym efektem, występuje bardzo często (Jackson, Jacobs, 1987; za: Jackson, O'Keefe, Jacobs, Brashers, 1989). W sytuacji wystąpienia takiej interakcji często obserwuje się układ wyników, w którym badany efekt zmiennej eksperymentalnej jest istotny statystycznie, gdy replikacje są czynnikiem stałym, ale staje się nieistotny statystycznie, gdy replikacje mają status czynnika losowego. Na gruncie badań psycholingwistycznych ilustrowali to Richter i Seay

(1987). Mowa jest tu więc o przypadku, w którym za istotny efekt zmiennej eksperymentalnej „odpowiedzialna" jest interakcja. W poprzednim akapicie przypadek ten oznaczono numerem 2. Jak przy okazji widać, jedynie wtedy, gdy replikacje uznajemy za czynnik losowy, jesteśmy w stanie dokładnie wskazać, co stoi za istotnym efektem zmiennej eksperymentalnej – czy zaszedł przypadek pierwszy, drugi, czy trzeci.

Wziąwszy pod uwagę przywołane już wyniki symulacji Monte Carlo, a także zasadne domniemanie, że replikacja „ma znaczenie", w tym sensie, że każda z nich do pewnego stopnia różni się efektem, należałoby wysunąć wątpliwość pod kątem wartości wyników przynajmniej części badań, w których replikacje traktowano jako czynnik stały i wyrazić obawę, że wynik może być *fałszywie dodatni* tam, gdzie odnotowano istotny efekt dla zmiennej eksperymentalnej.

A zatem pod względem sprawowania kontroli nad błędem I rodzaju, pożądanymi schematami eksperymentalnymi będą te, które replikacje traktują jako czynnik losowy, jako że pozwalają one utrzymać ten błąd na założonym poziomie *alfa* (Jackson, Brashers 1994b). Taki sam wniosek sformułowali Kenny i Judd (1986: 428), pisząc, że remedium na brak niezależności pomiarów jest uznanie czynnika grupującego obserwacje za losowy i włączenie go do analiz.

Na treści przedstawione w ostatnich akapitach można też spojrzeć przez pryzmat mocy testu, czyli zdolności testu do wykrycia efektu. Pod tym względem schematy krzyżowe z czynnikiem losowym wypadają mniej korzystnie, jeśli porównać je z tymi, w których replikacje traktuje się jako czynnik stały. Wracając do zmiennych uwzględnionych w przykładzie – jest całkiem prawdopodobne, że pojawiłby się taki układ wyników, w którym wystąpiłby istotny efekt metody nauczania, gdyby nauczyciele uznani byliby za czynnik stały, ale efekt ten „zniknąłby", gdyby nauczycieli uznać za czynnik losowy. Ujmuje się niekiedy tę sytuację jako pewien niefortunny paradoks – włączenie do modelu czynnika losowego to „sprowadzenie problemu", którego kosztu nie będzie ponosić czynnik będący jego źródłem. Obarczony będzie nim niestety czynnik stały odgrywający wiodącą rolę w hipotezach badawczych. Może to sprzyjać wrażeniu, że włączając czynnik losowy do modelu, badacz za swoją uczciwość i rzetelność jest raczej karany, a nie nagradzany, bo zmniejsza szansę wykrycia efektu istotnego statystycznie dla czynnika mającego dla niego dużą wagę (Howell, 2007: 434; Glass, Hopkins, 1996: 557). Rzeczywiście – co będzie dalej pokazane – efekty stałe odpowiadające wiodącym problemom badawczym, testowane są inaczej, gdy czynnik losowy jest obecny w modelu, niż w sytuacji gdyby

tego czynnika nie było. Kwestia ta dotyczy liczby stopni swobody. Jeśli czynnik losowy obecny jest w modelu, to liczba stopni swobody w mianowniku ilorazu F zwykle nie jest duża, bo jest pochodną liczby replikacji. Tym samym warunki do wykazania istotności efektu stałego nie są zbyt sprzyjające. Gdy czynnik losowy nie występuje w badaniu, liczba stopni swobody pochodzi od liczby osób badanych, która w praktyce zwykle jest na tyle duża, że tworzy dogodne warunki do wykazania efektu (przy założeniu, że efekt istnieje).

W dyskusji na temat mocy testów pojawia się też argument, że mówienie o modelu z czynnikiem losowym jako mającym mniejszą moc w porównaniu z modelem, w którym występują tylko czynnikami stałe, jest po części niewłaściwe, gdyż nie jest tak, że hipoteza zerowa dotycząca kluczowego czynnika stałego jest taka sama w obu modelach (Jackson, 1992: 112). Jak już o tym wspomniano wcześniej – w modelu z czynnikiem losowym hipoteza ta charakteryzuje się większym stopniem generalizacji, tj. wykraczającym poza przebadane replikacje, dlatego w tej sytuacji nie powinno dziwić, że wykazanie, iż głoszona w hipotezie prawidłowość występuje, może być trudniejsze, czy – mówiąc precyzyjniej – wiąże się z większymi wymogami. Będzie tu działać ta sama zasada, która dotyczy osób badanych – przy założeniu, że efekt istnieje, trudniej go będzie wykazać na małej niż na dużej próbie badawczej. Taką samą barierę będzie stwarzać niewielka liczebnie próba replikacji.

Potwierdzają to studia metodologiczne prowadzone dla schematu typu *T przez r*, czyli schematu *zabieg przez replikacje,* poświęcone wyznacznikom mocy dla kluczowego czynnika stałego (czynnika T, odpowiadającemu głównej manipulacji eksperymentalnej) (Jackson, Brashers, 1994a: 61, Jackson, Brashers, 1994b). Ich wyniki pokazują, po pierwsze, tę oczywistość, że moc testu uzależniona jest od wielkości efektu czynnika T – im silniejszy efekt (na przykład: im większe różnice między metodami nauczania), tym łatwiej go wykryć. Po drugie, efekt tym trudniej wykryć, im bardziej zmienia się on w zależności od replikacji (na przykład: efekt metody nie jest taki sam dla wszystkich nauczycieli, występuje interakcja między metodą nauczania a nauczycielem). Mówiąc krótko, tym trudniej jest wykryć efekt T, im większy jest efekt interakcji $T \times r$. Po trzecie, tym łatwiej jest wykryć efekt, im większa jest próba replikacji. Po czwarte, zakładając tę samą liczbę replikacji, tym łatwiej jest wykryć efekt T, im większa jest liczba osób w grupie.

Ilustracje tych prawidłowości wraz z obliczonymi wielkościami mocy, z uwzględnieniem wielkości efektu czynnika T (przyjmującego dwa poziomy: t_1 i t_2), interakcji $T \times r$, liczby replikacji przy liczebności

próby $n = 200$ oraz $n = 400$, Czytelnik znajdzie w załączniku 1. aneksu (tabele 27–28).

Załącznik 1. zawiera dodatkowo informację o mocy testu dla samej interakcji $T \times r$. Jak widać w zamieszczonych tam tabelach (27–28), zwiększenie liczby replikacji przy stałej liczebności próby uczestników (a więc zwiększając liczbę replikacji kosztem liczebności pojedynczej grupy eksperymentalnej) prowadzi do obniżenia zdolności wykrycia istotnej interakcji. O ile więc badaczowi zależy na dysponowaniu wysoką mocą dla testu interakcji, powinien on dysponować zarówno odpowiednio dużo liczbą replikacji, jak i odpowiednią liczbą jednostek badanych w pojedynczej grupie eksperymentalnej (w konsekwencji odpowiednio dużą próbą uczestników). Ilustruje to tabela 29., w której Czytelnik znajdzie wyniki analizy mocy testu dla interakcji $T \times r$ przy stałej liczebności pojedynczej grupy uczestników. Zestawienie to może być przydatne badaczom szukającym rozwiązania satysfakcjonującego, ale jednocześnie kompromisowego pod względem liczby replikacji i liczebności próby[11].

1.2.3.4. Ocena planu 2a

Przejdźmy teraz do omówienia ostatniego typu z czterech wyróżnionych schematów. Jest to schemat 2a, będący korektą i rozszerzeniem schematu 1a. Rozwiązanie 2a będzie zatem również schematem z replikacjami, choć jego konstrukcja będzie inna niż konstrukcja omówionego przed chwilą schematu 2b. Tym razem bowiem każdy nauczyciel uczy tylko jedną metodą, a więc jest „przyporządkowany" do m_1 albo do m_2. Mówiąc nieco bardziej formalnie, jedna część poziomów czynnika „nauczyciel" występuje na poziomie m_1, a druga część na poziomie m_2. Tym samym czynnik „nauczyciel" i czynnik „metoda" tworzą **strukturę zagnieżdżoną**[12] – czynnik „nauczyciel" jest zagnieżdżony w czynniku „metoda" (zob. rysunek 2.). Ujmując to w ogólniejszych kategoriach – replikacje są zagnieżdżone w zabiegach, stąd schemat ten można nazwać ***replikacje w zabiegu*** i w skrócie oznaczyć jako $\boldsymbol{r(T)}$.

[11] Test dla interakcji $T \times r$ powinien charakteryzować się odpowiednio dużą mocą, gdyby badacz rozważałby wariant, w którym w ostatecznym modelu pominie czynnik replikacje, o ile we wstępnych analizach udałoby się wykazać, że efekt interakcji jest nieistotny statystycznie. Jak argumentują Brashers i Jackson (1999: 464–465) tym praktykom niestety rzadko kiedy towarzyszy odpowiednia kontrola mocy testu dla interakcji. Zobacz też podrozdział 2.8.

[12] Strukturę zagnieżdżoną nazywa się też hierarchiczną. Podobnie nazywa się modele z taką strukturą.

Wadą schematu ze strukturą zagnieżdżoną jest konieczność użycia większej liczby replikacji (zaangażowania większej liczby nauczycieli), by uzyskać tę samą liczbę pomiarów, którą daje schemat ze strukturą krzyżową. Ekonomiczność schematów krzyżowych staje się tym większa, im więcej poziomów ma czynnik odnoszący się do zabiegu (tu: im więcej metod nauczania uwzględniono by w badaniu). Jednakże – o czym już wspomniano – nie każdy eksperyment jest wykonalny w schemacie krzyżowym. Sztandarowym przykładem schematów planowanych w strukturze zagnieżdżonej są te, w których zabiegiem badawczym jest płeć, czy to eksperymentatora, czy osoby badanej.

Konsekwencją omawianego układu czynników jest mniejsza liczba źródeł zmienności w porównaniu ze schematem 2b. Teraz możliwe jest wydzielenie zmienności pochodzącej od: 1) metody; 2) nauczycieli w obrębie poszczególnych metod 3) osób badanych, przy czym ta zmienność jest połączona ze zmiennością pochodzącą od niekontrolowanych czynników. Zatem warto zauważyć, że podczas gdy schemat 2a pozwala na wyodrębnienie tylko trzech źródeł zmienności (co jest jego istotnym ograniczeniem), schemat 2b pozwalał wyłonić cztery źródła zmienności.

Doprecyzowując, schemat, w którym dwa czynniki pozostają ze sobą w strukturze zagnieżdżonej, nie pozwala na wyodrębnienie zmienności pochodzącej z interakcji tych czynników – skoro m_1 występuje na innych poziomach czynnika „nauczyciel" niż m_2, to nie można ustalić, czy różnica między m_1 i m_2 (efekt metody) jest taka sama na każdym poziomie czynnika „nauczyciel" (dla każdego nauczyciela).

Ale układ ten ma też konsekwencje dla czynnika „nauczyciel" (replikacje). Schemat krzyżowy, w którym każdy z poziomów czynnika „nauczyciel" występował na każdym poziomie czynnika „metoda", pozwalał na „uniezależnienie" czynnika „nauczyciel" od czynnika „metoda". W schemacie zagnieżdżonym nie jest to możliwe – zmienność pochodzącą od nauczycieli można ustalić jedynie w obrębie poszczególnych metod, tj. oddzielnie dla m_1 i oddzielnie dla m_2 (dodając te zmienności do siebie, uzyskujemy wielkość zbiorczą efektu[13]). Istotny efekt nauczyciela w tym wypadku oznaczałby, że istotne różnice między nauczycielami obserwowane są w obrębie m_1 i/lub w obrębie m_2 (Maxwell, Delaney, Kelly, 2018: 574)[14].

[13] Kontynuację wątku i bliższe objaśnienie Czytelnik znajdzie w podrozdziale 2.7.

[14] Skoro porównywane są ze sobą schematy 2a i 2b, to warto także uzupełnić, że zmienność pochodząca od czynnika „nauczyciel" zagnieżdżonego w czynniku „metoda" zawiera w sobie połączone dwie zmienności: tę pochodzą od nauczycieli oraz tę pochodzącą od interakcji czynnika „nauczyciel" z czynnikiem „metoda", a więc zmienności, na których oddzielenie od siebie pozwalał schemat 2b ze strukturą krzyżową (Keppel, Wickens 2004: 558): Dokładniejsze objaśnienie Czytelnik znajdzie w podrozdziale 2.7.

Spróbujmy teraz – podobnie jak wcześniej – przeanalizować trzy warianty działań wobec czynnika „replikacje".

Pierwszy wariant polegałby na tym, by użyć replikacji (zaangażować co najmniej kilku nauczycieli) na etapie zbierania danych, jednak w analizach nie uwzględniać tego faktu, a więc nie włączać czynnika „replikacje". Obiekcje wobec takiego postępowania w odniesieniu do schematów ze strukturą zagnieżdżoną są tej samej natury, co wobec schematów ze strukturą krzyżową. Zignorować fakt, że replikacje (nauczyciele) „grupują" obserwacje w obrębie poszczególnych kategorii zmiennej eksperymentalnej (metod nauczania), to ryzykować tym, że obserwacje przestaną być niezależne. Podobieństwo zgrupowanych pomiarów wiąże się z inflacją błędu I rodzaju. Remedium na ten problem jest włączenie „zmiennej grupującej" (replikacje) do analiz w postaci czynnika losowego (Kenny, Judd, 1986: 427).

Drugi rozważany wariant to włączenie czynnika „replikacje" do analiz jako czynnika stałego. Linia argumentacyjna przeciw takiemu rozwiązaniu jest zasadniczo taka sama, jak w przypadku schematu ze strukturą krzyżową. Jeżeli replikacje traktowane są jako czynnik stały, to istnieje ryzyko, że hipoteza zerowa dla czynnika T, czyli czynnika odpowiadającego zmiennej eksperymentalnej, będzie testowana przy zawyżonym prawdopodobieństwie popełnienia błędu I rodzaju. Wskazują na to przeprowadzone symulacje Monte Carlo (Forster, Dickinson, 1976; Santa, Miller, Shaw, 1979; Wickens, Keppel, 1983)[15]. Specyfika omawianego obecnie schematu będzie natomiast miała to do siebie, że ryzyko inflacji błędu I rodzaju będzie zwiększać się tym bardziej, im bardziej zwiększać się będzie zmienność pochodząca od replikacji. Korzystając z przykładu – tym łatwiej będzie „wykryć" efekt metody (faktycznie nie istniejący), im bardziej różnić się będą od siebie nauczyciele wewnątrz metod, do których są przypisani. Przypomnijmy – przy rozwiązaniu ze strukturą krzyżową, inflacji błędu I rodzaju sprzyjała sytuacja, w której efekt metody różnił się w zależności od nauczyciela (inflacji błędu I rodzaju sprzyjała duża zmienność pochodząca od interakcji $T \times r$). W rozwiązaniu ze strukturą zagnieżdżoną nie da się ustalić efektu metody dla nauczyciela, ponieważ każdy z nich uczy tylko jedną metodą (nie da się ocenić interakcji $T \times r$), natomiast tym co podnosi prawdopodobieństwo wystąpienia błędu I rodzaju, czyli

[15] Schemat eksperymentu we wszystkich tych symulacjach uwzględniał taką strukturę, w której replikacje są zagnieżdżone w zmiennej zabiegowej, czyli $r(T)$, niemniej schemat ten był bardziej rozbudowany niż omawiany obecnie, jako że przewidywał poddanie badanych jednostek pomiarom wielokrotnym. Z tego względu szczegółowe ustalenia tych badań nie będą tu przytaczane.

sprzyja „wykryciu" efektu metody, jest wystąpienie efektu nauczyciela (wewnątrz metody), a więc efektu $r(T)$.

A zatem istotny statystycznie efekt zmiennej eksperymentalnej T w przypadku traktowania replikacji jako czynnika stałego, znów można objaśnić na trzy różne, wykluczające się sposoby: 1) efekt dla zmiennej eksperymentalnej T jest różny od zera oraz efekt replikacji jest równy zero 2) efekt dla zmiennej eksperymentalnej T jest równy zero, ale efekt replikacji jest różny od zera; 3) obydwa efekty są różne od zera (por. Maxwell, Delaney, Kelly, 2018: 576).

Dla uzmysłowienia wagi problemów, które stwarzają niepoprawne praktyki badawcze polegające na traktowaniu replikacji jako efektów stałych zamiast losowych, podejmowano powtórne analizy danych opublikowanych już badań. W obszarze psycholingwistycznym takie studia podejmował Clark (1973), który po włączeniu replikacji jako czynnika losowego wykazywał nieprawdziwość wcześniejszego wniosku o wystąpieniu efektu zmiennej eksperymentalnej. Takie re-analizy – z podobnym skutkiem – podejmowano też wobec badań z dziedziny komunikacji społecznej (Jackson, Brashers, Massey, 1992: 216). Tym samym sposobem Bonge, Schuldt i Harper (1992) zakwestionowali efekt płci eksperymentatora. Podobnie Dijkstra (1983) unaoczniał zniknięcie efektu płci ankietera, a także stylu prowadzenia wywiadu, w sytuacji, w której efekt samego ankietera zarysował się za mocno. W końcu też w badaniach Pilkonisa i in. (1984 za: Maxwell, Delaney, Kelly, 2018: 548) okazało się, że za efektem metody psychoterapii stoi psychoterapeuta, a nie metoda.

Do wyników tych rewizji należy przyłożyć odpowiednią wagę zważywszy na to, że objęto nimi pojedyncze studia empiryczne. Pamiętać przy tym należy, że to czy efekt zmiennej eksperymentalnej zostanie podtrzymany, gdy replikacje zostaną włączone jako czynnik losowy, zależy po części od liczby użytych replikacji i od liczebności próby. Nie można wykluczyć, że duża część eksperymentów, gdyby została poddana takim powtórnym analizom, okazałaby się poniekąd niedostosowana do takich testów, właśnie ze względu na nieodpowiednio małą liczbę replikacji czy liczbę osób w grupie. Wracamy tym samym do problemu mocy testu, czyli jego zdolności wykrycia istniejącego efektu.

W załączniku 2. aneksu Czytelnik znajdzie tabelę zawierającą wielkości mocy dla testu czynnika T odpowiadającego zmiennej eksperymentalnej. Tabele 30–31 unaoczniają, że efekt ten tym łatwiej jest wykryć 1) im jest większy; 2) im mniejszy jest efekt replikacji obserwowany na poziomach czynnika T, czyli im mniejszy efekt $r(T)$; 3) im większa jest liczba replikacji (nawet wtedy, gdy odbywa się to kosztem liczebności pojedynczej grupy); 4) im większa liczebność pojedynczej grupy, pod wa-

runkiem, że liczba replikacji nie zmienia się. Tabele 30–31 obrazują dodatkowo od czego zależy moc efektu replikacji, czyli efektu $r(T)$ – zwróćmy uwagę, że moc spada, gdy zwiększa się liczba replikacji kosztem liczebności grupy. Wyższej mocy sprzyja większa liczba replikacji, o ile nie wiąże się to ze zmniejszeniem liczebności grupy (tabela 32).

Kwestia siły efektu replikacji, a więc wielkości efektu ankietera czy terapeuty itd., jest też ważna sama w sobie. Gdyby wielkość ta miała okazać się znacząca, to warto byłoby ustalić, co się kryje za tym efektem. Skoro efekt ankietera czy terapeuty to efekt wiązki cech, dobrze byłoby wiedzieć, jaka cecha(y) z tej wiązki – inna niż płeć (ankietera), czy metoda terapii (stosowana przez terapeutę) – jest odpowiedzialna za to, że jedni ankieterzy czy terapeuci uzyskują wysokie, a inni niskie wartości zmiennej zależnej (Dijkstra, 1983).

Dodajmy również uwagę dotyczącą sposobu wybrania próby replikacji. Jak już pisano, pożądanym rozwiązaniem byłoby, aby replikacje dla poszczególnych kategorii zmiennej zabiegowej były wybrane na drodze losowej. Jeżeli jednak taki dobór nie jest możliwy, nie stanowi to przesłanki, by replikacje traktować jako efekt stały. O ile jako badacze chcemy uwolnić wnioski, zamiast ograniczać je do przebadanych replikacji – a trudno, żeby było inaczej – to powinniśmy traktować replikacje jakby stanowiły próbkę większej populacji replikacji i dopuścić, by efekt replikacji był bardziej zmienny niż efekt stały replikacji. Naturalnie przyjęcie takiego rozwiązania nie oznacza, że jest ono zupełnie nieproblematyczne. Wymaga ono przede wszystkim przemyślanego sposobu doboru próby, tak aby nie była próbą obciążoną. Nie jest bowiem tak, że uchylenie warunku próby losowej oznacza, że jakość próby nie ma znaczenia i że każda próba jest równie dobra. Zagadnienie to zostanie podjęte w kolejnym podrozdziale.

1.3. Nieprobabilistyczny dobór próby replikacji

Argumenty za zastosowaniem zabiegu randomizacji[16] w eksperymentach przedstawili w zwięzły sposób Winer, Brown i Michels (1991: 101 za: Gamst, Meyers, Guarino, 2008: 51):

> *Naruszenie wymogu losowego pobrania próby elementów z populacji i losowego przypisania elementów do warunków eksperymentalnych może całkowicie zdyskredytować badanie, ponieważ losowość zapewnia, że błędy będą niezależnie rozłożone wewnątrz grup*

[16] Zob. paragraf 1.2.1.

eksperymentalnych oraz między nimi. Losowość jest też mechanizmem wyeliminowania błędu systematycznego, jakim mogłyby być obciążone grupy eksperymentalne.

Zabieg randomizacji jest więc kluczowym warunkiem spełnienia jednego z najważniejszych założeń ANOVA[17]. Z tego punktu widzenia postulat Clarka (1973), który głosi, że należy dokonać rewizji podejścia w odniesieniu do tego, kiedy czynnik może być potraktowany jako losowy, a także by stosować wnioskowanie statystyczne wobec replikacji także wtedy, gdy nie zostały wyłonione losowo, został odebrany jako uderzający w fundamentalną zasadę. Zwłaszcza reakcje statystyków, w tym również tych o największym światowym uznaniu, były negatywne – broniąc pryncypium, zajęli stanowisko „*random means random*" (Wike, Church, 1976; Cohen, 1976; Keppel, 1976). Oponenci Clarka dodatkowo wskazywali, że 1) niektóre schematy z czynnikami losowym wymagają analizy za pomocą testów, które są tylko przybliżeniem testu F[18], a te nie są jeszcze wystarczająco dobrze zbadane; 2) obecność czynników losowych zwiększa błąd II rodzaju; oraz 3) modele z czynnikami losowymi wymagają mocnych założeń. Zgodni też byli co do tego, że w sytuacji, w której nie można pobrać losowej próby replikacji, czynnik replikacje należy potraktować jako stały, a uogólnienia wniosku na temat efektu szukać na drodze powtarzania eksperymentu. Powołując się na koncepcję Keppela o różnych rodzajach generalizacji, Wike i Church (1976) dodatkowo wskazali, że w osiąganiu ogólności wniosków, oprócz uzasadnień statystycznych, do dyspozycji badaczy są jeszcze uzasadnienia niestatystyczne. Była to więc rekomendacja konserwatywnej ścieżki działania, której poprawność Clark zakwestionował.

Podkreślić trzeba, że we wczesnej argumentacji oponentów nie pojawiły się głosy bezpośrednio odnoszące się do problemu, który podniósł Clark. Dopiero Wickens i Keppel (1983), wprawdzie zachowawczo podchodząc do proponowanego przez Clarka remedium, pisali o ewentualnym problemie dostarczania wyników fałszywie pozytywnych przez schematy, w których replikacje traktowane są jako czynnik stały (i jednocześnie sformułowane wnioski wykraczają poza przebadane replikacje). Natomiast dla Clarka był to problem o centralnym znaczeniu, który wzmacniał argumentem, że praktyki czasopism nakierowane są na publikowanie jedynie wyników istotnych statystycznie (1976) i że jest to kolejny mechanizm sprzyjający temu, by do puli twierdzeń naukowych wchodziły takie, które głoszą efekt, choć w rzeczywistości go

[17] Zob. podrozdział 2.4.
[18] Zob. podrozdział 2.3.

nie ma. W tym kontekście rada Wike'a i Churcha (1976), by badacze próbowali na drodze pozastatystycznych, a więc nierygorystycznych uzasadnień szukać racji dla możliwości dokonywania uogólnień, wydaje się jedynie pogłębiać problem.

Część argumentów wysuwanych przeciw propozycji Clarka została osłabiona przez wyniki metodologicznych badań przeprowadzonych w późniejszych czasie (i relacjonowanych w różnych fragmentach książki). Jak starałam się wcześniej pokazać, idee promowane przez Clarka spotkały się z pozytywnym oddźwiękiem w istotnej części środowiska naukowego, które za większy problem uznało praktyki prowadzenia badań z podwyższonym ryzykiem popełnienia błędu I rodzaju, niż odstępstwo od zasady, że za losowy można uznać czynnik tylko wtedy, gdy jego poziomy są losowo wybrane (np. Richter, Say, 1987). Sztywne trzymanie się pryncypium jawiło się tym bardziej jako postawa sztuczna, że praktyki doboru badanych jednostek były już dalekie od ortodoksji. Jak pisał Hopkins (1983: 107):

> *Niektórzy badacze błędnie konkludują, że skoro nauczyciele nie zostali wybrani losowo z dokładnie zdefiniowanej populacji nauczycieli, to nie powinni być traktowani jako czynnik losowy. A czy do badań eksperymentalnych, quasi-eksperymentalnych, czy korelacyjnych, studenci są rzeczywiście losowo wybierani z takiej dokładnie zdefiniowanej populacji studentów? Praktycznie nigdy; w rzeczywistości jest to zasadniczo niemożliwe bez zubożenia definicji populacji.*

Zdaniem Hopkinsa w praktyce badawczej punktem wyjścia jest raczej myślenie o próbie, a nie o populacji, w efekcie czego to definicja próby określa definicję populacji (1983: 108). I tu widać, że niekiedy zderzenie teorii z praktyką może urealnić oczekiwania i osłabić siłę zarzutów. Wike i Church (1976) pytali retorycznie, do jakiej populacji mają być odniesione wnioski na podstawie nielosowej próby replikacji? Pytanie to należy uznać za pozostające w mocy, także w odniesieniu do wielu prób badanych jednostek.

Zwolennicy podejścia Clarka, przede wszystkim Sally Jackson i jej współpracownicy, działali w obszarze nauk o komunikacji społecznej. Ideę traktowania replikacji jako czynnika losowego próbowali uprawomocnić w odniesieniu do replikacji mających postać komunikatów. Obszar ten ma swoją specyfikę, dlatego nie wszystkie tezy i propozycje sformułowane na tym gruncie daje się przełożyć na próby złożone z innych elementów niż komunikaty, np. na próby współpracowników badacza.

O specyfice tej świadczy już fakt, że skonstruowanie losowej próby komunikatów jest z gruntu niewykonalne, ponieważ nie można stwo-

rzyć dla nich operatu losowania. O ile w grę nie wchodzą przekazy już istniejące, np. zastosowane w reklamach, to komunikatów potencjalnych, dopiero dających się stworzyć jest nieskończenie wiele. Nie da się ich zatem wylistować, ponumerować ani określić dla nich prawdopodobieństwa dostania się do próby. Tak daleko idące ograniczenia nie dotykają operatów osób, choć i na tym polu badacze będą mieć często nieprzezwyciężalne trudności, by dysponować operatem jakieś wąskiej kategorii osób, np. nauczycieli czy ankieterów, itd. Niezależnie od natury tych barier przekreślają one możliwość skonstruowania próby losowej. Pryncypialne podejście do tej sytuacji wykluczałoby możliwość dokonywania jakichkolwiek generalizacji. Możliwości takiej pozbawione byłyby całe obszary niektórych dziedzin nauki. Jest to podejście – uważają Jackson i Jacobs (1983: 175) – zupełnie nie do przyjęcia.

Dlatego w tych obszarach, w których powstaje dylemat – uchylić się od generalizacji, czy generalizować przy podwyższonym prawdopodobieństwie błędu I rodzaju – należy poddać ponownej konceptualizacji relację między populacją a próbą. Dla tradycyjnego ujęcia charakterystyczna jest koncepcja populacji istniejącej przed próbą i niezależnie od próby, a także składającej się z wyraźnie określonych i wyodrębnionych elementów. W uproszczeniu – jest to myślenie od populacji do próby. Dla nowego ujęcia charakterystyczne byłoby myślenie o wektorze przeciwnym: od próby do populacji. Cechy elementów próby określałyby cechy elementów tworzących populację, limitując w ten sposób klasę zjawisk czy osób, do których miałyby odnosić się uogólnienia. Podkreślić należy, iż nie twierdzi się przy tym, że nowe ujęcie ma unieważnić tradycyjne, ani też, że obydwa ujęcia są równowartościowe. Oczywistym niedostatkiem nowego podejścia jest to, że pozostawia dużą dozę niepewności co do wartości, jaką przedstawia sobą populacja, do której mają odnosić się wnioski i silnie uzależnia tę wartość od procedury doboru, która zostanie przez badacza zastosowana. Nie pozwala też na wyraźne wyznaczenie granic tej populacji i wiąże się z dużą niepewnością odnośnie jakości szacunków. Jest to propozycja niewątpliwie radykalna, szczególnie jeżeli spojrzeć na nią jako na jawną próbę uprawomocnienia podejścia, które tradycyjne i podręcznikowe podejście wywraca do góry nogami. Natomiast radykalizm tej propozycji blednie, jeżeli wziąć pod uwagę, że w istocie odpowiada ona stosowanym powszechnie praktykom wobec prób badanych jednostek (zob. Jackson, Jacobs, 1983: 175–176).

O ile – jak wspomniano wyżej – dobór losowy jest wyposażony w mechanizm, który zasadniczo uwalnia próbę od błędów systematycznych (przynajmniej na etapie losowania), o tyle dobór nieprobabilistyczny otwiera drogę dla tego zagrożenia, stąd tak ważna jest jakość

próby celowej i przejrzystość procedur stosowanych przy jej konstruowaniu. Pewne sugestie dotyczące zasad doboru sformułowali Jackson i Jacobs (1983) oraz Jackson (1992). Zakładając, że konstruowana jest próba komunikatów, postulowali oni, by próbę tworzyły:

1) komunikaty prototypowe dla kategorii, którą mają reprezentować. Komunikaty nietypowe dla kategorii lub budzące wątpliwości, czy do niej należą powinny zostać odrzucone.
2) komunikaty, które pomyślnie przeszły pierwszy etap, powinny być maksymalnie zróżnicowane (np. ze względu na formę, treść, itd.). Taka strategia ma zabezpieczyć przed sytuacją, w której wynik badania będzie odzwierciedlać wybrany, pojedynczy wzór, będący pochodną podobieństwa komunikatów. Jest to przy okazji strategia, która sprzyjać będzie dużej wariancji błędu. Ostatecznie jednak takie działanie jest traktowane jako mniejsze zło niż niedoszacowanie tej wariancji.
3) Im więcej sztucznych ograniczeń badacz będzie nakładał na komunikaty, tym większe prawdopodobieństwo, że jego wizja „odpowiednich" komunikatów obciąży próbę błędem systematycznym. Z tego względu lepsza będzie próba komunikatów odznaczających się dużą „naturalnością", a więc głównie już istniejących, natomiast najgorsza to taka, która powstała w wyniku tworzenia i modyfikowania przez badacza komunikatu po komunikacie w celu dopasowania ich do hipotezy (z tego powodu niebezpieczeństwo mogą stwarzać badania wstępne, których wyniki mogą rodzić pokusę cyzelowania treści komunikatów w celu maksymalizacji efektu).

*

W podsumowaniu dotychczasowych rozważań wskażmy w syntetyczny sposób, jakim celom mogą służyć czynniki losowe w planach eksperymentalnych przewidujących obecność czynnika stałego, który to czynnik będzie mieć wiodącą rolę w hipotezach badawczych. Zasadniczo cele te można sprowadzić do trzech najważniejszych (i niewykluczających się): 1) eliminowanie czynników zakłócających, 2) zwiększenie zakresu stosowalności wniosków oraz 3) ocena rozmiaru zmienności towarzyszącej manipulacji badawczej (Jackson, Brashers, 1994a: 2). Gdyby odnieść te cele do przypadku modelu zawierającego jeden czynnik stały T, to można by je ująć jako – odpowiednio – zwiększenie trafności wewnętrznej wniosku dotyczącego efektu T, zwiększenie trafności zewnętrznej wniosku dotyczącego efektu T oraz ocena tego, na ile efekt T jest zmienny i zależny od replikacji. W przeciwieństwie do schematów bez replikacji, cele te z powodzeniem będą realizować schematy z repli-

kacjami, pod warunkiem oczywiście, że replikacje będą miały w analizach status czynnika losowego.

Czy koniecznym warunkiem realizacji tych celów jest także wylosowanie próby replikacji z populacji replikacji? Jak przedstawiłam to powyższym w tekście, początkowo zdania na ten temat były mocno podzielone, obecnie temperatura sporu zmalała, a stanowisko liberalne ma znacznie większą reprezentację w środowisku naukowym niż kiedyś. W ramach tego stanowiska co najmniej dopuszcza się użycie celowej próby replikacji. Takie podejście bliskie jest też mojemu. Uważam przy tym, że w sytuacji posłużenia się próbą celową, roszczenie do trafności zewnętrznej wniosków należy wysuwać z dużą ostrożnością i tylko wtedy, gdy istnieją mocne argumenty za tym, że próba nie jest obciążona. Sądzę też, że potencjału próby celowej w zapewnieniu eksperymentowi wysokiej trafności zewnętrznej nie powinno się – co do zasady – upatrywać w rozszerzenia zakresu wniosków na populację, gdyż prawomocność takiego rozszerzania zawsze pozostanie wątpliwa. Podniesienia trafności zewnętrznej przez próbę celową należy raczej upatrywać w tym, że w znaczący sposób chroni ona przed wyprowadzaniem wniosków o istnieniu efektu, podczas gdy w rzeczywistości tego efektu nie ma. Z tego punktu widzenia posłużenie się celową próbą replikacji – gdy losowa nie jest dostępna – nie jest już działaniem zaledwie „dopuszczonym", lecz działaniem wręcz pożądanym.

ROZDZIAŁ II

Analizy statystyczne

Analiza modeli dla schematów międzygrupowych (grup niezależnych) zawierających jeden czy dwa czynniki stałe jest stosunkowo łatwa. Odpowiadające tym modelom metody, odpowiednio jednoczynnikowa i dwuczynnikowa analiza wariancji z efektami stałymi, są omawiane jako pierwsze w każdym podręczniku poświęconym analizie wariancji i zwykle są obecne w każdym podręczniku stanowiącym wprowadzenie do statystyki. Wprawdzie modele dwu- i więcej czynnikowe są już koncepcyjnie i interpretacyjnie bardziej zaawansowane od modelu jednoczynnikowego, niemniej od strony obliczeniowej wiele zasad pozostaje wspólnych. Od strony analitycznej sytuacja zaczyna się komplikować, gdy do modelu wprowadzany jest dodatkowy, inny niż „osoby badane" czynnik losowy, lub gdy model dotyczy badania prowadzonego w schemacie wewnątrzgrupowym albo mieszanym[1]. W modelach tych przestaje obowiązywać zasada, że składnik błędu (mianownik w ilorazie F) jest ten sam dla wszystkich testowanych efektów i że zawsze będzie nim wariancja wewnątrzgrupowa. W modelach tych składnik błędu jest zawsze dobierany indywidualnie, tj. do konkretnego efektu oraz z uwzględnieniem zastosowanego schematu eksperymentu. Takiego indywidualnego dopasowania będą wymagały też analizy mające na celu oszacowanie wielkości poszczególnych efektów (ang. *effect size*) oraz analizy mocy testów dla poszczególnych efektów. Niezbędnym punktem wyjścia tych bardziej skomplikowanych analiz jest umiejętność wyznaczania *wartości oczekiwanych średnich kwadratów* (ang. *expected mean squares*), $E(MS)$.

---

[1] Nazwa schemat/model mieszany jest w literaturze używana w dwóch znaczeniach. Przy pierwszym znaczeniu chodzi o schemat, w którym badane jednostki poddawane są jednokrotnemu pomiarowi ze względu na jedną cechę oraz wielokrotnemu pomiarowi ze względu na drugą cechę. Schemat w tym sensie jest „mieszany", że zawiera elementy schematu międzygrupowego oraz schematu wewnątrzgrupowego (np. Keppel, Wickens 2004). W drugim znaczeniu, schemat jest „mieszany", ponieważ zawiera czynnik(i) stały oraz czynnik(i) losowe (poza czynnikiem „osoby badane") (np. Glass, Hopkins 1996, Sahai, Ageel 2000). Nazwę „model/schemat mieszany" rezerwuję dla pierwszego znaczenia. Tam, gdzie w grę wchodzi znaczenie drugie będę posługiwać się określeniem „model z efektami mieszanymi" (ang. *mixed effects ANOVA*).

W rozdziale II książki Czytelnik będzie mógł zapoznać się z metodą generowania $E(MS)$, a także dobierania odpowiedniego składnika błędu w konstruowaniu statystyki F (lub quasi-F) – przedstawione w tym zakresie procedury mają zastosowanie do wszystkich typów schematów badawczych, także wewnątrzgrupowych i mieszanych.

Dalej Czytelnik będzie miał możliwość zapoznania się z procedurami szacowania wielkości efektów i ustalania mocy testów – prezentację tych procedur ograniczam do schematów międzygrupowych, ale uwzględniam schematy z efektami stałymi, schematy z efektami losowymi i schematy z efektami mieszanymi (różnicując przy tym schematy, w których czynniki tworzą strukturę krzyżową lub zagnieżdżoną). Wszystkie przedstawione procedury zakładają dodatkowo, że grupy są równoliczne.

Ważną kwestią jest, czy komputerowe pakiety statystyczne mogą zwolnić badaczy analizujących dane uzyskane za pomocą złożonych schematów (np. z efektami losowymi, czy mieszanymi) z konieczności zaznajomienia się z metodami analizy, w tym ze wzorami. Trzeba mieć świadomość, że zdając się całkowicie na oprogramowanie badacz ryzykuje, że analizy nie zostaną wykonane poprawnie. Nawet korzystając z bardziej zaawansowanych procedur pakietu, pożądana jest ciągła kontrola poprawności wykonywanych analiz i gotowość do interwencji (np. Keppel, Wickens, 2004: 537, 554; Schwarz, 1993). Użytkownik pakietu statystycznego powinien raczej nastawić się na pisanie poleceń (kodów), samodzielne definiowanie zarówno składników modelu, jak i potrzebnych testów F, a niekiedy na dokonywanie obliczeń we własnym zakresie w oparciu o wzory.

Punktem wyjścia dla właściwego wywodu będzie skrótowe przypomnienie podstawowych idei z teorii analizy wariancji w odniesieniu do modelu jedno- i dwuczynnikowego z efektami stałymi dla grup niezależnych.

2.1. Podstawy analizy wariancji – modele z efektami stałymi w schemacie międzygrupowym

Rozważmy na początek **model jednoczynnikowy**, w którym T jest czynnikiem stałym, inaczej model jednoczynnikowy z efektem stałym T (ang. *fixed effect model*)[2]. W modelu tym czynnik T ma t poziomów, a więc utworzył t grup, a w każdej z grup jest s obserwacji (osób badanych). Wartość zmiennej zależnej dla pojedynczej osoby będzie oznaczona jako y_{ij} (wynik j-tej obserwacji należącej do i-tej grupy).

[2] Nazwa ang. *fixed effect model* w polskim piśmiennictwie jest też tłumaczona jako model o ustalonych efektach (np. Aczel 2000: 424, Rószkiewicz i in. 2013).

Punktem wyjścia analizy jest model statystyczny, rozpatrywany na poziomie populacji, który opisuje, w jaki sposób kształtuje się y_{ij}. Gdyby „zdekomponować" wynik pojedynczej osoby, to składałby się on z następujących trzech elementów:

$$y_{ij} = \mu + (\mu_i - \mu) + \left(y_{ij} - \mu_i\right)$$

i w skrótowej postaci:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

gdzie:
μ to średnia ogólna badanej cechy w populacji (jej wartość nie jest znana);
$\alpha_i = \mu_i - \mu$ to różnica między średnią populacyjną w i-tej grupie (jej wartość nie jest znana) a ogólną średnią populacyjną – efekt i-tego poziomu czynnika T;
$\epsilon_{ij} = y_{ij} - \mu_i$ to różnica między wynikiem j-tej obserwacji w i-tej grupie a średnią populacyjną w i-tej grupie – łączny efekt wpływu innych czynników niż T (Rozmus 2011: 133; Sahai, Ageel, 2000: 11), inaczej **błąd eksperymentalny** dla j-tej obserwacji w i-tej grupie (Howell, 2007: 464; Keppel, 1982: 82; Brzeziński, 2008: 92), a mówiąc bardziej uniwersalnym językiem analizy regresji, jest to składnik losowy, reszty modelu (Aczel, 2000: 402, 459; Glass, Hopkins, 1996: 388).

Ten sam model strukturalny, tym razem na poziomie próby (w oparciu o dane zebrane w badaniu), wyglądałby następująco (Brzeziński, 2008: 92–93):

$$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + \left(y_{ij} - \bar{y}_i\right)$$

Jest więc y_{ij} sumą następujących komponentów[3]:
$\bar{y}$ – średniej ogólnej badanej cechy w próbie;
$(\bar{y}_i - \bar{y})$ – różnicy między średnią z i-tej grupy a średnią ogólną;
$\left(y_{ij} - \bar{y}_i\right)$ – różnicy między wynikiem j-tej obserwacji należącej do i-tej grupy a średnią z i-tej grupy.

[3] Przykładowo, jeżeli zmienną zależną byłby wzrost, czynnikiem T byłaby płeć, średnia wzrostu w badanej próbie wynosiła $\bar{y} = 170$, średnia wzrostu w grupie mężczyzn wynosiła $\bar{y}_i = 177$, a wzrost konkretnego badanego mężczyzny wynosił $y_{ij} = 172$, to wynik y_{ij} można przedstawić jako $172 = 170 + (177 - 170) + (172 - 177)$.

Odejmując $\bar{y}$ od obu stron równania, uzyskujemy wyrażenie pozwalające zdekomponować już nie pojedynczy wynik, ale całkowite odchylenie pojedynczego wyniku od średniej ogólnej:

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

Wyrażenie to jest „bazową" konstrukcją, która jest wykorzystywana przy syntetyzowaniu wyników uzyskanych od wszystkich uczestników badania. Całkowitą zmienność wyników ujmiemy jako:

$$SS_{total} = \sum_i \sum_j (y_{ij} - \bar{y})^2$$

A następnie zdekomponujemy tę wielkość według wcześniejszej zasady:

$$\sum_i \sum_j (y_{ij} - \bar{y})^2 = \sum_i \sum_j (\bar{y}_i - \bar{y})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_i)^2$$

co w kategoriach ogólnych, można przedstawić jako:

$$SS_{total} = SS_{between-groups} + SS_{within-groups}$$

gdzie:
SS_{total} – całkowita suma kwadratów odchyleń od średniej (ang. *total sum of squares*), zmienność całkowita;
$SS_{between-groups} = \sum_i \sum_j (\bar{y}_i - \bar{y})^2 = s \sum_i (\bar{y}_i - \bar{y})^2$ – międzygrupowa suma kwadratów odchyleń od średniej (ang. *between-groups sum of squares*), zmienność międzygrupowa;
$SS_{within-groups}$ – wewnątrzgrupowa suma kwadratów odchyleń od średniej (ang. *within-groups sum of squares*), zmienność wewnątrzgrupowa.

Zmienność międzygrupowa wynika z działania badanego czynnika, a więc tutaj zmienność wyjaśniona przez T. Możemy to zapisać jako:

$$SS_{between-groups} = SS_T$$

Zmienność wewnątrzgrupowa to zmienność niewyjaśniona przez T, czyli taka, którą należy przypisać oddziaływaniu innych czynników niż T.

A mówiąc jeszcze dokładniej, jest to zróżnicowanie, które pozostaje między badanymi jednostkami po pogrupowaniu ich przez czynnik T. Skoro więc mielibyśmy badane jednostki ująć jako czynnik, to moglibyśmy przypisać mu literę s (od ang. *subjects*). Ostateczny zapis tego czynnika musiałby jeszcze uwzględniać to, że każda z osób badanych przypisana jest do jednej grupy utworzonej przez czynnik T (do jednego poziomu czynnika T), dlatego czynnik losowy „osoby badane" powinien zostać ostatecznie ujęty jako $s(T)$. A zatem:

$$SS_{within-groups} = SS_{s(T)}$$

Próbując wyrazić to jeszcze dokładniej, $SS_{s(T)}$ odzwierciedla zróżnicowanie, które utrzymuje się u badanych osób pomimo faktu, że w ramach danej grupy eksperymentalnej mają takie same warunki. Zróżnicowanie to wynika, po pierwsze, z oddziaływania trwałych, indywidualnych charakterystyk osób badanych, którymi te osoby różnią się między sobą (jednak w wyniku zastosowania zabiegu randomizacji ich oddziaływanie można uznać za efekt losowy), a po drugie, z oddziaływania innych niekontrolowanych losowych źródeł, stanowiących błąd w węższym sensie – mogą nimi być błędy pomiaru, chwilowe zmiany zachodzące u osób badanych i dotyczące poziomu ich uwagi, motywacji, itd. (Keppel, Wickens, 2004: 353, 19–20). Cechą charakterystyczną omawianego schematu, podobnie jak każdego schematu międzygrupowego, jest to, że zmienności pochodzącej od osób badanych nie da się uzyskać w czystej postaci, tj. wyizolować od oddziaływań drugiego typu. Także samego błędu rozumianego w węższym sensie nie da się wyizolować za pomocą żadnego dającego się wyobrazić schematu eksperymentalnego – swoim oddziaływaniem zawsze będzie on zakłócać bądź to czynnik (tak jak tu: czynnik „osoby badane"), bądź interakcję czynników, jak to się dzieje w schematach wewnątrzgrupowych i mieszanych (Jackson, Brashers, 1994a: 63)[4].

Dekompozycję całkowitej zmienności możemy też zatem przedstawić jako:

$$SS_{total} = SS_T + SS_{s(T)}$$

[4] Zapis, który w pełni oddawałby tę złożoność, mógłby mieć w przypadku omawianego planu następującą postać: $SS_{s(T),error}$. Taką rozszerzoną notację, uwzględniającą błąd w wąskim znaczeniu i jego nakładanie się na inne konkretne źródła, stosują badacze uprawiający *generalizabliity theory* Cronbach i in. (1972), Shavelson i Webb (1991), Brennan (2001). Specyfiką ich podejścia jest to, że w równaniu modelu strukturalnego nie uwzględniają reszt modelu (ϵ) (Cronbach i in. 1972: 1–2), nie przeprowadzają też testów istotności.

Sprawdzając, czy efekt czynnika T jest istotny statystycznie, należy wcześniej uzyskać dwie wielkości – wariancję międzygrupową oraz wariancję wewnątrzgrupową. Przez wariancję rozumiemy – zasadniczo – przeciętne odchylenie kwadratowe wyników od średniej. Jeżeli wyniki opisują próbę, to nieco bardziej formalny zapis mówiłby, że (Keppel, Wickens, 2004: 24, 32):

$$wariancja = \frac{suma\ kwadratów\ odchyleń\ od\ średniej}{liczba\ stopni\ swobody} = \frac{SS}{df} = MS$$

W metodzie analizy wariancji wielkość ta nazywana jest również średnim kwadratem odchyleń od średniej, czy jeszcze krócej: średnim kwadratem (ang. *mean square*) i oznaczana jako MS.

Wariancję międzygrupową, a więc średni kwadrat dla czynnika T, uzyskamy dzięki formule (Keppel, Wickens, 2004: 32–33):

$$MS_{between-groups} = MS_T = SS_T/df_T \qquad df_T = t - 1$$

Z kolei wariancję wewnątrzgrupową, a więc średni kwadrat dla czynnika $s(T)$ uzyskamy w następujący sposób:

$$MS_{within-groups} = MS_{s(T)} = SS_{s(T)}/df_{s(T)} \quad df_{s(T)} = t(s-1)$$

W następnym kroku analiz obliczana jest wartość statystyki testowej, która pozwoli ustosunkować się do hipotezy zerowej. Testując efekt czynnika T, hipotezę zerową oraz alternatywną można sformułować jako (por. Keppel, Wickens, 2004: 134):

$$H_0: wszystkie\ \mu_i\ są\ równe, czyli\ \mu_1 = \mu_2 = \mu_3 = etc.$$

$$H_1: nieprawda, że\ wszystkie\ \mu_i\ są\ równe\ (przynajmniej\ dwie\ \mu_i\ nie\ są\ równe).$$

Równoważnie do powyższych zapisów, hipotezy te można sformułować bezpośrednio w kategoriach efektu i-tego poziomu czynnika (tutaj czynnika T):

$$H_0: wszystkie\ \alpha_i\ są\ równe\ zero, czyli\ \alpha_1 = 0\ , \alpha_2 = 0\ , \alpha_3 = 0\ , etc.$$

$$H_1: nieprawda, że\ wszystkie\ \alpha_i\ są\ równe\ zero.$$

Lub w jeszcze bardziej skrótowej formie:

$$H_0: \sum \alpha_i^2 = 0$$

$$H_1: \sum \alpha_i^2 \neq 0$$

W teście głównego efektu T statystyką testową będzie F – iloraz dwóch wielkości: wariancji międzygrupowej i wewnątrzgrupowej:

$$F_{(df_T,\, df_{s(T)})} = \frac{MS_{between-groups}}{MS_{within-groups}} = \frac{MS_T}{MS_{s(T)}}$$

Wartość statystyki F odnosimy do centralnego rozkładu F o $df_T = t - 1$ i $df_{s(T)} = t(s-1)$ stopniach swobody. Gdy wartość F jest wystarczająco duża i wpada do obszaru krytycznego (czy – analizując sprawę z drugiej strony – gdy wartość p jest mniejsza od ustalonego poziomu istotności $alfa$), odrzucamy H_0 na rzecz H_1 i konkludujemy, że efekt czynnika T jest istotny statystycznie, zatem przynajmniej między niektórymi średnimi grupowymi zachodzi różnica istotna statystycznie (jest za duża by można ją było przypisać losowej zmienności).

Rozważmy teraz model **dwuczynnikowy** z efektami stałymi T oraz R w strukturze krzyżowej. Tym razem oprócz czynnika stałego T o t poziomach do modelu wejdzie czynnik stały R o r poziomach. Ponieważ T i R są w strukturze krzyżowej, to liczba grup wynosi $t \times r$. Jak wcześniej, w każdej grupie jest s osób badanych.

Na poziomie populacji model strukturalny przybiera postać:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

gdzie:
y_{ijk} jest k-tą obserwacją na i-tym poziomie czynnika T oraz na y-tym poziomie czynnika R,
$\alpha_i = \mu_i - \mu$ to efekt i-tego poziomu czynnika T,
$\beta_j = \mu_j - \mu$ to efekt j-tego poziomu czynnika R,
$(\alpha\beta)_{ij} = \mu_{ij} - \mu_i - \mu_j + \mu$ to efekt interakcji czynników T i R (efekt interakcji i-tego poziomu czynnika T z j-tym poziomem czynnika R),
ϵ_{ijk} jest błędem losowym związanym z k-tą obserwacją i-tego poziomu czynnika T i j-tego poziomu czynnika R.

Całkowita suma kwadratów odchyleń od średniej jest znów dekomponowana na zmienność międzygrupową oraz wewnątrzgrupową:

$$SS_{total} = SS_{between-groups} + SS_{within-groups}$$

Zmienność międzygrupowa to zmienność wyjaśniona – składa się na nią zmienność wynikająca z działania czynnika T, zmienność wynikająca z działania czynnika R oraz zmienność wynikająca z działania interakcji czynników T i R[5].

$$SS_{between-groups} = SS_T + SS_R + SS_{T\times R}$$

Każda z tych trzech porcji zmienności wyjaśnionej jest wyznaczana oddzielnie:

$$SS_T = \sum_i \sum_j \sum_k (\bar{y}_i - \bar{y})^2 = rs \sum_i (\bar{y}_i - \bar{y})^2$$

$$SS_R = \sum_i \sum_j \sum_k \left(\bar{y}_j - \bar{y}\right)^2 = ts \sum_j \left(\bar{y}_j - \bar{y}\right)^2$$

$$SS_{T\times R} = \sum_i \sum_j \sum_k \left[\bar{y}_{ij} - (\bar{y}_i - \bar{y}) - \left(\bar{y}_j - \bar{y}\right) - \bar{y}\right]^2$$

$$= s \sum_i \sum_j \left(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}\right)^2$$

Zmienność wewnątrzgrupowa to zmienność niewyjaśniona – zróżnicowanie, które utrzymuje się u badanych osób w grupach wydzielonych przez tworzące strukturę krzyżową czynniki T i R. Przypisanie zmienności niewyjaśnionej do osób badanych oddaje zapis:

$$SS_{within-groups} = SS_{s(T\times R)}$$

[5] Podział na dwa efekty główne i interakcję jest konwencjonalnym sposobem podziału zmienności międzygrupowej. Na marginesie można dodać, że w tym modelu zmienność międzygrupowa może być zdekomponowana na dwa dodatkowe sposoby (Keppel, Wickens 2004: 253):

$$SS_{between-groups} = \sum SS_{T\ na\ poziomie\ r_j} + SS_R$$

$$SS_{between-groups} = \sum SS_{R\ na\ poziomie\ t_i} + SS_T$$

Zmienność ta wyznaczana jest według wzoru:

$$SS_{s(T\times R)} = \sum_i \sum_j \sum_k \left(\bar{y}_{ijk} - \bar{y}_{ij}\right)^2$$

Wielkości średnich kwadratów dla poszczególnych źródeł zmienności otrzymamy według wzorów:

$$MS_T = SS_T/df_T \qquad df_T = t - 1$$
$$MS_R = SS_R/df_R \qquad df_R = r - 1$$
$$MS_{T\times R} = SS_{T\times R}/df_{T\times R} \qquad df_{T\times R} = (t-1)\times(r-1)$$
$$MS_{s(T\times R)} = SS_{s(T\times R)}/df_{s(T\times R)} \qquad df_{s(T\times R)} = (s-1)\times t\times r$$

Hipoteza zerowa w teście głównego efektu czynnika T przyjmuje postać:

$$H_0\colon \sum \alpha_i^2 = 0 \text{ (versus } H_1\colon \sum \alpha_i^2 \neq 0)$$

Hipotezy zerowa w teście głównego efektu czynnika R przyjmuje postać:

$$H_0\colon \sum \beta_j^2 = 0 \text{ (versus } H_1\colon \sum \beta_j^2 \neq 0)$$

Hipotezy zerowa w teście głównego efektu interakcji T i R przyjmuje postać:

$$H_0\colon \sum (\alpha\beta)_{ij}^2 = 0 \text{ (versus } H_1\colon \sum (\alpha\beta)_{ij}^2 \neq 0)$$

Do testu głównego efektu czynnika T będzie wykorzystana statystyka:

$$F_{(df_T,\, df_{s(R\times T)})} = \frac{MS_T}{MS_{within-groups}} = \frac{MS_T}{MS_{s(R\times T)}}$$

Do testu głównego efektu czynnika R będzie wykorzystana statystyka:

$$F_{(df_R,\, df_{s(T\times R)})} = \frac{MS_R}{MS_{within-groups}} = \frac{MS_R}{MS_{s(T\times R)}}$$

Do testu głównego efektu interakcji czynników T i R będzie wykorzystana statystyka:

$$F_{(df_{T\times R},\, df_{s(T\times R)})} = \frac{MS_{T\times R}}{MS_{within-groups}} = \frac{MS_{T\times R}}{MS_{s(T\times R)}}$$

Jak widać, statystyka F dla wszystkich dotąd rozpatrywanych efektów była konstruowana w ten sam sposób – jako iloraz wariancji międzygrupowej (lub pewnej jej części) i wariancji wewnątrzgrupowej. Zgodnie z tą zasadą w mianowniku ilorazu F występowała wariancja wewnątrzgrupowa i miało to miejsce zarówno wtedy, gdy czynnik T był jedynym czynnikiem wyjaśniającym, jak i wtedy, gdy w modelu pojawił się drugi czynnik R. Zasada ta była adekwatna dlatego, że zarówno czynnik T, jak i czynnik R były czynnikami stałymi. Tymczasem, gdyby drugi czynnik wprowadzony do modelu był czynnikiem losowym, to konstrukcja statystyki F dla czynnika T byłaby inna, a zmiana dotyczyłaby mianownika – właściwym **składnikiem błędu** (ang. *error term*) nie byłaby już wariancja wewnątrzgrupowa $MS_{s(T\times R)}$. Akurat w tym przypadku odpowiednim składnikiem błędu byłaby $MS_{T\times r}$. Dodajmy jeszcze, że podobny problem dotyczy każdego schematu wewnątrzgrupowego i mieszanego – one również wymagają indywidualnego dobrania składnika błędu dla każdego testowanego efektu.

Zasada uniwersalna, która rządzi konstrukcją statystyki F niezależnie od zastosowanego schematu i charakteru obecnych w nim czynników (losowy vs. stały), opiera się na *oczekiwanych* średnich kwadratach odchyleń od średnich $E(MS)$. **Ogólna zasada mówi, że iloraz F mają tworzyć dwa średnie kwadraty MS, których wartości oczekiwane $E(MS)$ różnią się tylko obecnością komponentu odnoszącego się do testowanego efektu. Składnikiem błędu ma być ten MS, którego wartość oczekiwana $E(MS)$ zawiera wszystkie komponenty oprócz komponentu dla efektu.**

Ideę można przedstawić za pomocą uproszczonych, konceptualnych formuł w następujący sposób (Keppel, Wickens, 2004: 226):

$$E\left(MS_{efekt}\right) = efekt + błąd\ dla\ efektu$$

$$E\left(MS_{błąd\ dla\ efektu}\right) = błąd\ dla\ efektu$$

a wtedy uogólnioną postać ilorazu F można ująć jako:

$$F = \frac{MS_{efekt}}{MS_{błąd\ dla\ efektu}}$$

Wróćmy do schematu jednoczynnikowego dla grup niezależnych z efektem stałym T. Oczekiwane średnie kwadraty dla kolejnych źródeł w tym schemacie przedstawia tabela 1[6]. Jak widać, $E(MS_T)$ oraz $E\left(MS_{s(T)}\right)$ różnią się tylko jednym elementem, tj. komponentem dla testowanego efektu T, widniejącym tu jako $s\theta_T^2$, zaś wspólnym elementem jest komponent błędu $\sigma_{s(T)}^2$. W świetle przedstawionej zasady, $MS_{s(T)}$ jest odpowiednim składnikiem błędu do ilorazu F, stąd właśnie

$$F_{(df_T,\, df_{s(T)})} = \frac{MS_T}{MS_{s(T)}}$$

Tabela 1. Oczekiwane średnie kwadraty dla schematu z efektem stałym T

Źródło	MS	Oczekiwane średnie kwadraty $E(MS)$	
T	MS_T	$s\theta_T^2$	$+ \sigma_{s(T)}^2$
$s(T)$	$MS_{s(T)}$		$\sigma_{s(T)}^2$

Źródło: opracowanie własne.

Z kolei tabela 2. przedstawia oczekiwane średnie kwadraty dla schematu dwuczynnikowego[7]. Do testu efektu T iloraz F powinien być

[6] Przypomnijmy, że omawianemu schematowi odpowiada model $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$. Zawarte w tabeli 1. komponenty $E(MS)$ to (Sahai, Ageel 2000: 26):

$$\theta_T^2 = \frac{\sum(\alpha_i)^2}{(t-1)} = \frac{\sum(\mu_i - \mu)^2}{(t-1)}$$

Jest więc θ_T^2 miarą zmienności (rozproszenia) grupowych średnich populacyjnych od ogólnej średniej populacyjnej. Hasłowo, choć niezbyt ściśle nazywana jest wariancją. Zważywszy na obecność parametru μ w liczniku, na miano to zasługiwałyby wtedy, gdyby w mianowniku widniało t, a nie $(t-1)$ (Howell, 2007: 323, 433; Kirk, 1968: 58). W wielu podręcznikach statystycznych jedynie w tekście odnotowuje się tę różnicę, ale na poziomie zapisu pomija się dystynkcję między θ^2 a σ^2, i stosuje się σ^2 (np. Winer, 1971: 163; Glass, Hopkins, 1996: 394).

$\sigma_{s(T)}^2$ czyli wariancja reszt ϵ_{ij} (Brown, Prescott, 2015: 4). Natomiast w nazewnictwie przyjętym w analizie komponentów wariancyjnych, $\sigma_{s(T)}^2$ będzie nosić miano komponentu wariancyjnego (ang. *variance component*), tj. liniowego składnika całkowitej wariancji σ_y^2. W omawianym modelu jednoczynnikowym z efektem stałym T, $\sigma_{s(T)}^2$ akurat wyczerpuje wariancję całkowitą (jest jej jedynym komponentem), co jest ujęte jako $\sigma_y^2 = \sigma_{s(T)}^2$ (Brown, Prescott, 2015: 8).

[7] Gdzie: $\sigma_{s(T\times R)}^2$ jest wariancją reszt ϵ_{ijk}. W ujęciu charakterystycznym dla analizy komponentów wariancyjnych byłby $\sigma_{s(T\times R)}^2$ komponentem wariancyjnym (zważywszy jednak

zbudowany jako $MS_T/MS_{s(T\times R)}$, gdyż oczekiwanym średnim kwadratem, który różni się od $E(MS_T)$ jedynie obecnością $rs\theta_T^2$ będzie $E\left(MS_{s(T\times R)}\right)$. Analogicznie zostaną przetestowane efekt czynnika R oraz efekt interakcji $T \times R$ – w każdym z tych testów właściwym składnikiem błędu będzie $MS_{s(T\times R)}$.

Tabela 2. Oczekiwane średnie kwadraty dla schematu z efektami stałymi w strukturze krzyżowej $T \times R$

Źródło	MS	Oczekiwane średnie kwadraty $E(MS)$			
T	MS_T	$rs\theta_T^2$			$+\ \sigma_{s(T\times R)}^2$
R	MS_R		$ts\theta_R^2$		$+\ \sigma_{s(T\times R)}^2$
$T \times R$	$MS_{T\times R}$			$s\theta_{T\times R}^2$	$+\ \sigma_{s(T\times R)}^2$
$s(T \times R)$	$MS_{s(T\times R)}$				$\sigma_{s(T\times R)}^2$

Źródło: opracowanie własne.

Przypomnijmy, że wartość oczekiwana danej statystyki jest średnią jej *rozkładu z próby* (ang. *sampling distribution*) (Glass, Hopkins, 1996: 393). Rozkład ten powstałby poprzez powtarzanie nieskończoną liczbę razy sekwencji czynności: pobieranie z populacji próby losowej o ustalonej liczebności n; przeprowadzanie eksperymentu według tych samych zasad; rejestrowanie wartości jaką dana statystyka osiągnęła w aktualnej próbie. Rozważając przypadek schematu jednoczynnikowego, interesującymi nas statystykami byłyby $MS_{between-groups} = MS_T$ oraz $MS_{within-groups} = MS_{s(T)}$. Dla każdej wylosowanej próby wielkości te byłyby odnotowywane po to, aby mógł powstać rozkład międzygrupowych średnich kwadratów oraz rozkład wewnątrzgrupowych średnich kwadratów. Średnia rozkładu międzygrupowych średnich kwadratów byłaby wartością oczekiwaną MS_T, a średnia rozkładu

na to, że w omawianym modelu jest jedynym efektem losowym, to $\sigma_{s(T\times R)}^2 = \sigma_y^2$), θ^2 dla poszczególnych efektów to (Sahai, Ageel 2000: 194, Maxwell, Delaney, Kelly 2018: 557):

$$\theta_T^2 = \frac{\sum(\alpha_i)^2}{(t-1)}$$

$$\theta_R^2 = \frac{\sum(\beta_j)^2}{(r-1)}$$

$$\theta_{T\times R}^2 = \frac{\sum\sum(\alpha\beta)_{ij}^2}{(t-1)(r-1)}$$

wewnątrzgrupowych średnich kwadratów byłaby wartością oczekiwaną $MS_{s(T)}$.

Praktyczna korzyść wynikająca ze znajomości komponentów składających się na $E(MS)$ dla konkretnych źródeł to – jak już wiadomo – dostarczenie podstawy do poprawnego skonstruowania statystyki F. Jaką inną, praktyczną informację dostajemy dowiadując się, że $E(MS_T) = s\theta_T^2 + \sigma_{s(T)}^2$ oraz $E\left(MS_{s(T)}\right) = \sigma_{s(T)}^2$? Mówiąc swobodnie, znając $E(MS_T)$ dowiadujemy się, od czego zależy zaobserwowana w próbie wielkość MS_T, i analogicznie – znając $E\left(MS_{s(T)}\right)$ dowiadujemy się, od czego zależna jest odnotowana w próbie wielkość $MS_{s(T)}$ (Keppel, Wickens, 2004: 135).

Wielkość $MS_{s(T)}$ jest w całości pochodną wielkości błędu eksperymentalnego ϵ_{ij}, na który będą się składać: błąd związany z losowaniem (lub rozlosowaniem), zróżnicowanie między badanymi osobami występujące pomimo tego, że mają w grupie te same warunki, błędy pomiaru, itd. Całość tej zmienności „lokowana" jest zbiorczo w badanych jednostkach i traktowana jako zmienność o losowym charakterze.

Natomiast od czego zależy wielkość MS_T, czyli co wpływa na takie, a nie inne wartości średnich grupowych w próbie $\bar{y}_i$? Jak widać, MS_T zależy od wielkości efektu T w populacji (od tego ile wynoszą μ_i), a dodatkowo od zmienności losowej. Zdanie to przynosi niezbędne uściślenie tego, co powiedziano wcześniej na temat zmienności międzygrupowej – MS_T odzwierciedla nie tylko sam efekt T, ale także błąd eksperymentalny. *De facto* nie ma w tym nic zaskakującego – skoro wyniki, którymi dysponujemy są wynikami próby, która jest tylko jedną z wielu innych prób możliwych do wylosowania z populacji, a dodatkowo wyniki te mogą być obarczone innymi błędami, np. pomiaru, to nie możemy oczekiwać, że $\bar{y}_1$ będzie równać się dokładnie μ_1, itd. Z tego samego powodu nie oczekujemy, że przypadek prawdziwości hipotezy zerowej, czyli sytuacji, w której $\mu_1 = \mu_2 = \mu_3 = etc$ ma prawo zrealizować się w jedynie poprzez taki układ wyników, w którym $\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = etc$. (Keppel, 1982: 52; Keppel, Wickens, 2004: 18, 20–21; Glass, Hopkins, 1996: 388–389).

Jak nadmieniono wcześniej, zmienność losowa przypisana do badanych osób nie musi być jedyną zmiennością losową, która „wpłynie" na średnie grupowe w próbie $\bar{y}_i$, dlatego $MS_{within-groups}$ nie zawsze będzie właściwym składnikiem błędu (mianownikiem) ilorazu F. Jeżeli na średnie grupowe w próbie $\bar{y}_i$ wpływałaby jakaś dodatkowa porcja losowej zmienności, to składnik błędu również powinien ją uwzględniać, tak aby uczynić zadość zasadzie mówiącej, że mianownik ma zawierać

wszystkie komponenty, które zawiera licznik poza komponentem dla efektu. Dlaczego respektowanie tej zasady jest takie ważne? Zauważmy, że jeżeli całość losowej zmienności znajdującej się w liczniku nie pojawi się w mianowniku, to wartość licznika będzie przeszacowana. To z kolei prowadzi do przeszacowania wartości F. A zatem nierespektowanie zasady oznacza tworzenie warunków, które w sposób nieuprawniony będą sprzyjały wykazaniu efektu.

W schematach złożonych – schematach wieloczynnikowych zawierających zarówno czynniki stałe, jak i losowe – liczba wszystkich dających się wyodrębnić źródeł zmienności może urosnąć do kilkunastu czy nawet dwudziestu kilku (na przykład Richter, Seay, 1987; Glass, Hopkins, 1996: 565), a oczekiwany średni kwadrat dla jednego źródła może składać z trzech i więcej komponentów. W takich schematach wylistowanie wszystkich losowych źródeł zmienności, które kształtują $\bar{y}_i$ byłoby bardzo trudne bez algorytmu, który pozwalałby je wskazywać. Algorytm, o którym mowa, to zasady wyznaczania oczekiwanych średnich kwadratów $E(MS)$.

2.2. Wyznaczanie oczekiwanych średnich kwadratów $E(MS)$

Wyznaczanie $E(MS)$ w oparciu o wzory matematyczne (na przykład Kirk, 1968: 53–58, 509–512) jest dość skomplikowane i wymaga przygotowania w tym zakresie, dlatego wielu autorów zaproponowało mechaniczne reguły generowania $E(MS)$ (na przykład Keppel, 1982: 638–641; Winer, 1971: 371–375; Hopkins, 1976; Jackson, Brashers, 1994a: 17–19; Glass, Hopkins, 1996: 555–556; Sahai, Ageel, 2000: 591–595). W prezentowanym opracowaniu przywołane zostaną zasady zaproponowane przez Jackson i Brashers (1994a) dlatego, że wydają się najprostsze w użyciu. O najważniejszych różnicach między powyższymi propozycjami informuję w przypisach dalszej części tekstu.

Reguły wyznaczania $E(MS)$ zostaną ujęte w trzech podstawowych krokach. Po zaprezentowaniu każdego kroku pokazuję sposób jego realizacji dla schematów 2a oraz 2b. Przypomnijmy, że każdy z tych schematów przewidywał, że w układzie badawczym wystąpią: manipulacja eksperymentalna (ang. ***t**reatments*), replikacje (ang. ***r**eplications*) oraz osoby badane (ang. ***s**ubjects*). Stosowane przeze mnie dalej oznaczenia biorą się od pierwszych liter tych angielskich nazw. Oznaczenia te wykorzystuję następnie do przedstawienia całego systemu zapisu, a system ten sam w sobie jest ważną częścią zasad generowania $E(MS)$. Jest on zresztą stosowany przeze mnie konse-

kwentnie w całym rozdziale II. Na wszelki wypadek dodam, że zaprezentowane zasady generowania $E(MS)$, wraz z systemem zapisu, nie stosują się tylko do omawianych tu schematów – wręcz przeciwnie, są uniwersalne, nadają się do rozpatrzenia każdego planu eksperymentalnego, oddając jego specyfikę[8].

Krok I. Lista źródeł zmienności (czynników i interakcji) w schemacie.

a) **Wylistuj wszystkie czynniki, włącznie z czynnikiem „badane osoby"**[9]. Każdemu czynnikowi przypisz literę alfabetu łacińskiego. Dla odróżnienia czynników stałych i losowych duża litera będzie przypisywana czynnikom stałym, a mała – losowym[10].

b) **Wyłoń wszystkie czynniki zagnieżdżone**. Jeżeli jakiś czynnik jest zagnieżdżony w innym czynniku, na przykład, jeśli czynnik r jest zagnieżdżony w czynniku T, to – zgodnie z przyjętą tu formą zapisu – zostanie to ujęte jako $r(T)$. Zapis ten oddaje relację między czynnikiem zagnieżdżonym, a zagnieżdżającym według schematu[11]: *czynnik zagnieżdżony* (*czynnik zagnieżdżający*).
Zapis czynnika, który jest zagnieżdżony, nigdy nie może pomijać faktu, że czynnik ten występuje w strukturze zagnieżdżonej. Jeżeli więc czynnik r jest zagnieżdżony w T, to błędem byłoby wyodrębnić zarówno r, jak i $r(T)$ – poprawne jest wyłonić jedynie $r(T)$.

[8] Niezależnie od przyjętego systemu generowania $E(MS)$ oddzielną grupę przypadków będą stanowić schematy „niekompletne". Chodzi o schematy, w których badacz świadomie decyduje się na zredukowanie danych, np. dokonuje pomiarów tylko dla części kombinacji poziomów czynników (np. zastosuje schemat kwadratu łacińskiego) lub gdy decyduje się na schemat międzygrupowy, w którym w każdej grupie eksperymentalnej będzie tylko jedna badana jednostka, a nie co najmniej dwie. Taka redukcja danych pociągnie za sobą niemożność wyodrębnienia niektórych źródeł zmienności.

[9] Tym samym przyjęte jest tu rozwiązanie stosowane przez Hopkinsa (1976), Keppela (1982), Keppela i Wickensa (2004), Jackson i Brashers (1994a), Gamsta, Meyers, Guarino (2008), polegające na tym, by zawsze – niezależnie od schematu – wyodrębniać czynnik „osoby badane". Przy konwencjonalnym rozwiązaniu (np. Winer 1971, Kirk 1968, Glass, Hopkins 1996, Howell 2007) czynnik „osoby badane" wyodrębnia się tylko w schematach wewnątrzgrupowych (dla grup zależnych) i mieszanych. Przy obecnym rozwiązaniu czynnik „osoby badane" będzie również wyodrębniany w schematach międzygrupowych (dla grup niezależnych).

[10] Tradycyjne rozwiązania (np. Winer 1971: 371) nie uwzględniają tej różnicy w zapisie.

[11] W podręcznikach można się spotkać z różnymi konwencjami zapisu struktury zagnieżdżonej. Stosowana tutaj konwencja przewiduje użycie okrągłych nawiasów do zamieszczania czynnika zagnieżdżającego, stąd zapis $r(T)$, podczas gdy w innych publikacjach będzie on umieszczany po ukośniku, wówczas zapis przyjmuje postać r/T czy po dwukropku, wtedy zapis przyjmuje postać $r{:}\,T$.

Często zdarza się, że dany czynnik jest zagnieżdżony w więcej niż jednym czynniku jednocześnie. Przykładowo czynnik „osoby badane" s może być zagnieżdżony w r i jednocześnie być zagnieżdżony w T. Dane źródło może być wyodrębnione tylko raz, a zatem niepoprawnie byłoby wyodrębnić badane osoby jako dwa źródła tj. $s(T)$ oraz $s(r)$. Poprawny zapis czynnika s powinien oddawać nie tylko to, że on sam znajduje się w strukturze hierarchicznej (zagnieżdżonej), ale także to, w jakiej relacji pozostają ze sobą r oraz T. Jeżeli więc r i T tworzyłyby strukturę hierarchiczną, taką że $r(T)$, to zapis czynnika s powinien przyjąć postać $s(r(T))$. Natomiast jeżeli r oraz T tworzyłyby strukturę krzyżową, taką że $r \times T$, to zapis czynnika s powinien wyglądać $s(r \times T)$[12].

Krok 1b musi poprzedzać krok 1c – poprawne określenie wszystkich czynników zagnieżdżonych jest warunkiem koniecznym poprawnego wyłonienia wszystkich dopuszczalnych interakcji.

c) **Wylistuj wszystkie dozwolone interakcje**. Na pierwszym etapie można utworzyć mechanicznie wszystkie potencjalne interakcje jako kombinacje pomiędzy wszystkimi czynnikami (kombinacje dwuskładnikowe, trzy- i więcej składnikowe, o ile jest tyle czynników). Na drugim etapie należy zostawić tylko te interakcje, które są faktycznie dozwolone, tj. wyeliminować wszystkie te, przy których obserwujemy, że ta sama litera znajduje się jednocześnie wewnątrz i na zewnątrz nawiasu.

Realizacja kroków I a–c dla schematu 2b.

Schemat 2b zawiera trzy czynniki, przy czym tylko badane osoby są czynnikiem zagnieżdżonym (realizacja kroków I a–b):

$$T$$

$$r$$

$$s(r \times T)$$

[12] W wielu podręcznikach stosuje się konwencję skróconego zapisu czynników tworzących komponent zagnieżdżający. Zgodnie z tą konwencją, zarówno $s(r(T))$, jak i $s(r \times T)$ zostałyby ujęte jako $s(rT)$, a więc z pominięciem informacji o tym, jaką strukturę tworzą elementy komponentu. W obecnym opracowaniu stosowany jest zapis pełny; głównie przez wzgląd na użytkowników pakietów statystycznych, którzy mogą stanąć przed zadaniem samodzielnej specyfikacji źródeł zmienności w modelu (włącznie z czynnikiem „osoby badane"), a później samodzielnego zdefiniowania testów (określenia postaci ilorazu F).

a zatem wszystkie możliwe interakcje, jakie możemy utworzyć na pierwszym etapie kroku I c, to będą:

$$r \times T$$
$$T \times s(r \times T)$$
$$r \times s(r \times T)$$
$$T \times r \times s(r \times T)$$

Stosując regułę obowiązującą na drugim etapie kroku I c, zachowamy tylko interakcję, która widnieje pierwsza na liście, czyli $r \times T$. Interakcję drugą na liście należy usunąć, gdyż litera T występuje wewnątrz i na zewnątrz nawiasów (wyrażenie musi być usunięte, gdyż nie jest możliwe, aby czynnik s był zagnieżdżony w czynniku T i jednocześnie wchodził z nim w interakcję). Interakcja trzecia na liście musi być wyeliminowana ze względu na występowanie litery r w obu miejscach (nie jest możliwe, aby czynnik s był zagnieżdżony w czynniku r i jednocześnie wchodził z nim w interakcję). Interakcja czwarta w kolejności zawiera w obu miejscach litery r oraz T, stąd również jest do usunięcia. A zatem źródła zmienności w schemacie 2b tworzyć będą trzy czynniki i jedna interakcja, a konkretnie:

$$T$$
$$r$$
$$r \times T$$
$$s(r \times T)$$

Realizacja kroków I a–c dla schematu 2a.
Schemat 2a zawiera następujące czynniki, z czego dwa są czynnikami zagnieżdżonymi (realizacja kroku I a–b):

$$T$$
$$r(T)$$
$$s(r(T))$$

Na pierwszym etapie kroku I c powstaną następujące interakcje:

$$T \times r(T)$$
$$T \times s(r(T))$$
$$r(T) \times s(r(T))$$
$$T \times r(T) \times s(r(T))$$

Niemniej żadna z nich nie przejdzie pomyślnie etapu drugiego. A zatem źródła zmienności w schemacie 2a tworzą same czynniki, nie ma żadnej interakcji:

$$T$$
$$r(T)$$
$$s(r(T))$$

Krok II. Lista składników wariancji.

W oparciu o listę źródeł zmienności (powstałą w kroku I) będzie tworzona lista składników wariancji. Proces ten obejmuje następujące czynności:

a) **Dla każdego czynnika (nie dla interakcji) ustal liczbę jego poziomów**. Liczbę poziomów zapisujemy za pomocą litery, tej samej którą oznaczony został czynnik, ale zawsze litery małej (niezależenie od tego, czy sam czynnik jest uznany za losowy, czy stały)[13].

b) **Do każdego źródła zmienności przyporządkuj wyrażenie σ^2 albo θ^2**. Wyrażenie σ^2 będzie świadczyć o tym, że źródło jest losowe, a wyrażenie θ^2 o tym, że źródło jest stałe. Wyrażenie σ^2, nazywane komponentem wariancyjnym (ang. *variance component*), należy przypisywać tylko do czynników losowych oraz do interakcji zawierających co najmniej jeden czynnik losowy. Wyrażenie σ^2 należałoby więc przypisać do takich przykładowych czynników jak: r, czy $s(t)$, ale także $s(T)$ oraz do takich przykładowych interakcji jak: $r \times t$, czy $r \times T$. Jak widać, obecność małej litery w nazwie źródła czyni je losowym. Wyrażenie θ^2 (oznaczające wielkość zróżnicowania efektów stałych, patrz przypis 6. i 7.), należy przypisywać jedynie do czynników stałych oraz do interakcji zawierających jedynie czynniki stałe. W indeksie dolnym wyrażenia należy umieścić nazwę źródła[14].

---

[13] Przy wprowadzonym rozwiązaniu, np. czynnik stały A miałby a poziomów, czynnik stały B miałby b poziomów, czynnik losowy c miałby c poziomów. W tradycyjnych rozwiązaniach (np. Winer 1971) rezerwuje się oddzielne symbole dla czynników i oddzielne symbole dla poziomów czynników, np. czynnik A miałby p poziomów, czynnik B miałby q poziomów, czynnik c miałby r poziomów.

[14] W tradycyjnym zapisie (np. Winer 1971, Jackson 1992, Sahai, Ageel 2000) w indeksie dolnym nie umieszcza się – tak jak tutaj – nazwy **źródła** (przy użyciu litery lub liter alfabetu łacińskiego, np. A), lecz nazwę **efektu dla tego źródła** na poziomie populacji (przy użyciu litery lub liter alfabetu greckiego, np. α). Trzeba odnotować, że tradycyjny zapis jest bardziej precyzyjny, a także tworzy bezpośrednie odwołanie do odpowiedniego wyrażenia w modelu strukturalnym. Zaletą przyjętego tutaj uproszczonego zapisu jest to, że znacznie ułatwia on wyznaczanie $E(MS)$.

c) **Przyporządkuj współczynniki do wyrażeń powstałych w kroku II b**. Aby z „zawiązki" składnika (rezultatu kroku II b) powstał cały składnik należy przypisać do niego współczynnik. Na pierwszym etapie współczynnik ten tworzy iloczyn liter wyłonionych w kroku II a. Na drugim etapie ze współczynnika należy usunąć wszystkie litery, które występują w indeksie dolnym, przy czym wielkość liter (mała vs. duża) nie ma znaczenia.

Realizacja kroków II a–c dla schematu 2b.
Czynnik T ma t poziomów (a gdyby uwzględnić plan konkretnego badania, np. plan badawczy pokazany na rysunku 2., to $t = 2$). Czynnik r ma r poziomów ($r = 5$). Czynnik „osoby badane", a więc czynnik zagnieżdżony $s(r \times T)$ będzie miał s poziomów, przy czym tutaj liczbę poziomów należy rozumieć jako liczbę osób badanych przypadających na jedną grupę eksperymentalną ($s = 5$). Czynniki T i r będące w strukturze krzyżowej $r \times T$ utworzyły łącznie 10 grup, osób badanych n jest łącznie 50, natomiast osób w jednej grupie jest $s = 5$. Jak widać, iloczyn poziomów poszczególnych czynników równa się liczebności próby, tzn. $t \times r \times s = n$.

Abstrahując od konkretnego planu badawczego, w schemacie 2b występują (realizacja kroku II a):

$$t$$
$$r$$
$$s$$

W wyniku realizacji kroku II b powstaną następujące „zawiązki" składników:

$$\theta_T^2$$
$$\sigma_r^2$$
$$\sigma_{r\times T}^2$$
$$\sigma_{s(r\times T)}^2$$

Współczynnik będący iloczynem liter wyłonionych w kroku II a jest równy trs. Czynności obejmujące pierwszy etap kroku II c przyniosą rezultat w postaci:

$$trs\theta_T^2$$
$$trs\sigma_r^2$$
$$trs\sigma_{r\times T}^2$$
$$trs\sigma_{s(r\times T)}^2$$

W ramach drugiego etapu dla każdej „zawiązki” dokonujemy rewizji liter zawartych we współczynniku trs i usuwamy te z nich, które dublują się z literami występującymi w indeksie dolnym:

~~t~~$rs\theta_T^2$ (ze współczynnika usuwamy literę t, gdyż T występuje w indeksie dolnym).

t~~r~~$s\sigma_r^2$ (ze współczynnika usuwamy r, gdyż r występuje w indeksie dolnym).

~~tr~~$s\sigma_{r\times T}^2$ (ze współczynnika usuwamy r oraz t, gdyż r i T występują w indeksie dolnym).

~~trs~~$\sigma_{s(r\times T)}^2$ (usuwamy cały współczynnik trs, gdyż litery s, r oraz T występują w indeksie dolnym).

Otrzymujemy ostateczną listę składników:

$$rs\theta_T^2$$

$$ts\sigma_r^2$$

$$s\sigma_{r\times T}^2$$

$$\sigma_{s(r\times T)}^2$$

Realizacja kroków II a–c dla schematu 2a.

Czynnik T ma t poziomów (a biorąc pod uwagę plan badawczy na rysunku 4., to $t = 2$). Czynnik zagnieżdżony $r(T)$ ma r poziomów, przy czym tutaj liczbę poziomów należy rozumieć jako liczbę poziomów czynnika r przypadających na jeden poziom czynnika T (czynnik T utworzył 2 grupy, więc choć wszystkich nauczycieli jest 10, to $r = 5$). Czynnik „osoby badane”, a więc zagnieżdżony $s(r(T))$ będzie miał s poziomów, przy czym tutaj liczbę poziomów należy rozumieć jako liczbę osób badanych w jednej grupie eksperymentalnej. Czynniki T i r będące w strukturze zagnieżdżonej $r(T)$ utworzyły łącznie 10 grup, osób badanych n jest łącznie 50, natomiast osób w jednej grupie jest $s = 5$. Jak widać, iloczyn poziomów poszczególnych czynników równa się liczebności próby, tzn. $t \times r \times s = n$.

Abstrahując od konkretnego planu badawczego, w schemacie 2a występują (realizacja kroku II a):

$$t$$

$$r$$

$$s$$

W wyniku realizacji kroku II b powstaną:

$$\theta_T^2$$

$$\sigma_{r(T)}^2$$

$$\sigma_{s(r(T))}^2$$

Współczynnik będący iloczynem liter wyłonionych w kroku II a, jest równy trs. Czynności obejmujące pierwszy etap kroku II c dadzą rezultat:

$$trs\theta_T^2$$

$$trs\sigma_{r(T)}^2$$

$$trs\sigma_{s(r(T))}^2$$

A w ramach drugiego etapu powstanie następująca ostateczna lista składników:

$$rs\theta_T^2$$

$$s\sigma_{r(T)}^2$$

$$\sigma_{s(r(T))}^2$$

Krok III. Wybór komponentów tworzących oczekiwany średni kwadrat dla poszczególnych źródeł.

a) **Zestaw kolejno każde źródło (wyłonione w kroku I) z pełną listą komponentów (wyłonioną w kroku II c)**.
b) **Do docelowej listy komponentów zakwalifikuj te, które w indeksie dolnym zawierają wszystkie litery występujące w nazwie źródła**. Warunek ten spełnią komponenty zawierające w indeksie dolnym te i tylko te litery, które występują w nazwie źródła, a także komponenty, które oprócz tego, że zawierają wszystkie litery źródła, zawierają także litery dodatkowe. Postępowanie wobec komponentów z nadwyżkowymi literami, tj. nie występującymi w nazwie źródła, reguluje poniższy krok III c.
c) **Usuń komponent z docelowej listy zawsze wtedy, gdy nadwyżkową literą w indeksie dolnym jest duża litera niebędąca w nawiasie.** Akceptowana jest obecność dużych liter pod warunkiem, że są w nawiasie. Zawsze akceptowana jest obecność małych liter.

d) **Dla każdego źródła utwórz sumę komponentów, które przetrwały selekcję przeprowadzoną w ramach kroków III b i III c.** Powstała suma stanowi oczekiwany średni kwadrat dla źródła.

Realizacja kroków III a–d dla schematu 2b.

Realizację podpunktów kroku III warto przeprowadzić z wykorzystaniem tabeli.

Zgodnie z zasadą określoną dla kroku III a, w tabeli 3. dla każdego źródła (wiersze kolumny „Krok I") przygotowano pełną listę komponentów wariancji (wiersze kolumny „Krok III a") – jest to lista wyjściowa do przeprowadzenia selekcji.

Z myślą o końcowej postaci tabeli, zamieszczono w niej dodatkowo ustalenia, które powstały w wyniku realizacji kroków II a oraz II c. Te ustalenia nie będą w tej chwili potrzebne, dlatego kolumny, które je zawierają, zostały oznaczone kolorem szarym.

Tabela 3. Realizacja kroku III a dla schematu 2b

Krok I	Krok II a	Krok II c	Krok III a			
Źródła	Poziomy	Komponenty				
T	t	$rs\theta_T^2$	$rs\theta_T^2$	$ts\sigma_r^2$	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
r	r	$ts\sigma_r^2$	$rs\theta_T^2$	$ts\sigma_r^2$	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
$r\times T$		$s\sigma_{r\times T}^2$	$rs\theta_T^2$	$ts\sigma_r^2$	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
$s(r\times T)$	s	$\sigma_{s(r\times T)}^2$	$rs\theta_T^2$	$ts\sigma_r^2$	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$

Źródło: opracowanie własne.

W ramach kroku III b przeglądamy wyjściową listę komponentów pod kątem liter znajdujących się w indeksach dolnych i zostawiamy tylko te komponenty, które zawierają wszystkie litery źródła – czynności te wykonujemy po wierszu, dla każdego źródła oddzielnie.

W przypadku czynnika T, usuwamy z listy komponent $ts\sigma_r^2$, gdyż jako jedyny nie zawiera w indeksie dolnym litery T. Na liście zostają pozostałe komponenty, gdyż w ich indeksach dolnych występuje litera T.

Na analogicznej zasadzie – w przypadku czynnika r z listy usuniemy komponent $rs\theta_T^2$, ponieważ w jego indeksie dolnym nie występuje litera r. Pozostałe komponenty mają r w indeksach dolnych, dlatego komponenty te zostają na liście.

Przeanalizujmy postępowanie wobec interakcji $r \times T$ – w tym przypadku musimy usunąć komponent $rs\theta_T^2$, jako że w jego indeksie nie występują litery T **oraz** r (brakuje r); z tego samego powodu musi być wykreślony $ts\sigma_r^2$ (brakuje T).

W przypadku źródła $s(r \times T)$ na liście komponentów zostanie tylko $\sigma_{s(r\times T)}^2$, ponieważ tylko w jego indeksie dolnym są wszystkie litery źródła.

Tabela 4. przedstawia rezultat czynności wykonanych w ramach kroku III b dla każdego źródła.

Tabela 4. Realizacja kroku III b dla schematu 2b

Krok I	Krok II a	Krok II c	Krok III b			
Źródła	Poziomy	Komponenty				
T	t	$rs\theta_T^2$	$rs\theta_T^2$	~~$ts\sigma_r^2$~~	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
r	r	$ts\sigma_r^2$	~~$rs\theta_T^2$~~	$ts\sigma_r^2$	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
$r \times T$		$s\sigma_{r\times T}^2$	~~$rs\theta_T^2$~~	~~$ts\sigma_r^2$~~	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
$s(r \times T)$	s	$\sigma_{s(r\times T)}^2$	~~$rs\theta_T^2$~~	~~$ts\sigma_r^2$~~	~~$s\sigma_{r\times T}^2$~~	$\sigma_{s(r\times T)}^2$

Źródło: opracowanie własne.

Dalszą selekcję przeprowadzamy przy użyciu reguły określonej w kroku III c. Reguła ta mówi, że z listy komponentów należy usunąć taki, w którego indeksie dolnym jest litera inna niż w nazwie źródła i jednocześnie jest to litera duża oraz niebędąca w nawiasie.

W wyniku zastosowania tej reguły, dla czynnika T zachowamy wszystkie dotychczasowe komponenty – zachowany zostanie $s\sigma_{r\times T}^2$, gdyż nadwyżkowa litera r jest literą małą; zachowany zostanie też $\sigma_{s(r\times T)}^2$, gdyż nadwyżkowe litery r oraz s są literami małymi.

W przypadku czynnika r, należy usunąć z listy komponent $s\sigma_{r\times T}^2$ z uwagi na wystąpienie w jego indeksie litery T, która jest literą nadwyżkową, dużą i niebędącą w nawiasie; komponent $\sigma_{s(r\times T)}^2$ należy jednak zachować na liście – litera T jest wprawdzie nadwyżkowa, ale znajduje się w nawiasie.

W przypadku interakcji $r \times T$ żaden dodatkowy komponent nie zostanie usunięty – nadmiarowa litera s występująca w komponencie $\sigma_{s(r\times T)}^2$ jest literą małą.

W przypadku źródła $s(r \times T)$ jedyny przysługujący mu komponent $\sigma_{s(r\times T)}^2$ nie zawiera liter nadmiarowych, dlatego w ogóle nie podlega rewizji.

Tabela 5. przedstawia rezultat czynności wykonanych w ramach kroku III c dla poszczególnych źródeł.

Tabela. 5. Realizacja kroku III c dla schematu 2b

Krok I	Krok II a	Krok II c	Krok III c			
Źródła	Poziomy	Komponenty				
T	t	$rs\theta_T^2$	$rs\theta_T^2$	~~$ts\sigma_r^2$~~	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
r	r	$ts\sigma_r^2$	~~$rs\theta_T^2$~~	$ts\sigma_r^2$	~~$s\sigma_{r\times T}^2$~~	$\sigma_{s(r\times T)}^2$
$r\times T$		$s\sigma_{r\times T}^2$	~~$rs\theta_T^2$~~	~~$ts\sigma_r^2$~~	$s\sigma_{r\times T}^2$	$\sigma_{s(r\times T)}^2$
$s(r\times T)$	s	$\sigma_{s(r\times T)}^2$	~~$rs\theta_T^2$~~	~~$ts\sigma_r^2$~~	~~$s\sigma_{r\times T}^2$~~	$\sigma_{s(r\times T)}^2$

Źródło: opracowanie własne.

Suma komponentów, które przetrwały selekcję, tworzy oczekiwany średni kwadrat dla źródła. Oczekiwane średnie kwadraty dla poszczególnych źródeł w schemacie 2b przedstawia tabela 6[15].

Tabela 6. Realizacja kroku III d dla schematu 2b. Oczekiwane średnie kwadraty dla schematu z efektem stałym T i efektem losowym r w strukturze krzyżowej $T\times r$

Krok I	Krok II a	Krok II c	Krok III d			
Źródła	Poziomy	Komponenty	Oczekiwane średnie kwadraty $E(MS)$			
T	t	$rs\theta_T^2$	$rs\theta_T^2$		$+\ s\sigma_{r\times T}^2$	$+\ \sigma_{s(r\times T)}^2$
r	r	$ts\sigma_r^2$		$ts\sigma_r^2$		$+\ \sigma_{s(r\times T)}^2$
$r\times T$		$s\sigma_{r\times T}^2$			$s\sigma_{r\times T}^2$	$+\ \sigma_{s(r\times T)}^2$
$s(r\times T)$	s	$\sigma_{s(r\times T)}^2$				$\sigma_{s(r\times T)}^2$

Źródło: opracowanie własne.

[15] Gdzie (Sahai, Ageel, 2000: 194):

$$\theta_T^2 = \frac{\sum(\alpha_i)^2}{(t-1)}$$

natomiast σ_r^2, $\sigma_{r\times T}^2$ oraz $\sigma_{s(r\times T)}^2$ to komponenty wariancyjne.

Realizacja kroków III a–d dla schematu 2a.
Dla każdego źródła zmienności występującego w schemacie 2a przygotowujemy wyjściową, pełną listę komponentów. Rezultat tych działań przedstawia tabela 7. (kolumna „Krok I" oraz kolumna „Krok III a").

Tabela 7. Realizacja kroku III a dla schematu 2a

Krok I	Krok II a	Krok II c	Krok III a		
Źródła	Poziomy	Komponenty			
T	t	$rs\theta_T^2$	$rs\theta_T^2$	$s\sigma_{r(T)}^2$	$\sigma_{s(r(T))}^2$
$r(T)$	r	$s\sigma_{r(T)}^2$	$rs\theta_T^2$	$s\sigma_{r(T)}^2$	$\sigma_{s(r(T))}^2$
$s(r(T))$	s	$\sigma_{s(r(T))}^2$	$rs\theta_T^2$	$s\sigma_{r(T)}^2$	$\sigma_{s(r(T))}^2$

Źródło: opracowanie własne.

Wyniki pierwszej selekcji komponentów przedstawia tabela 8. Dla czynnika T zachowywane są wszystkie wyjściowe komponenty, gdyż każdy z nich zawiera w indeksie dolnym literę T. Z kolei dla czynnika $r(T)$ zachowywane są $s\sigma_{r(T)}^2$ i $\sigma_{s(r(T))}^2$, gdyż zawierają w swoich indeksach dolnych zarówno r, jak i T, natomiast komponent $rs\theta_T^2$ jest usuwany z listy jako niespełniający tego warunku. W przypadku czynnika $s(r(T))$ zachowany został tylko komponent $\sigma_{s(r(T))}^2$, gdyż jako jedyny zawiera w indeksie dolnym wszystkie trzy litery źródła: s, r, T.

Tabela 8. Realizacja kroku III b dla schematu 2a

Krok I	Krok II a	Krok II c	Krok III b		
Źródła	Poziomy	Komponenty			
T	t	$rs\theta_T^2$	$rs\theta_T^2$	$s\sigma_{r(T)}^2$	$\sigma_{s(r(T))}^2$
$r(T)$	r	$s\sigma_{r(T)}^2$	~~$rs\theta_T^2$~~	$s\sigma_{r(T)}^2$	$\sigma_{s(r(T))}^2$
$s(r(T))$	s	$\sigma_{s(r(T))}^2$	~~$rs\theta_T^2$~~	~~$s\sigma_{r(T)}^2$~~	$\sigma_{s(r(T))}^2$

Źródło: opracowanie własne.

Wyniki drugiej selekcji komponentów przedstawia tabela 9. Przypomnijmy, że w ramach tej selekcji podejmowana jest decyzja dotycząca komponentów zawierających litery „nadwyżkowe". Jak widać z tabeli, tym razem żaden komponent nie kwalifikował się do usunięcia. W przypadku źródła T wszystkie litery „nadwyżkowe" (inne niż T)

występujące w indeksach komponentów są literami małymi. Podobnie rzecz się ma w odniesieniu do czynnika $r(T)$. Literą nadmiarową w przypadku tego źródła jest litera s występująca w komponencie $\sigma^2_{s(r(T))}$, ale ponieważ jest literą małą, komponent musi zostać zachowany. Z kolei komponent dla czynnika $s(r(T))$, czyli komponent $\sigma^2_{s(r(T))}$, w ogóle nie zawiera liter nadmiarowych i dlatego nie podlega rewizji.

Tabela 9. Realizacja kroku III c dla schematu 2a

Krok I	Krok II a	Krok II c	Krok III c		
Źródła	Poziomy	Komponenty			
T	t	$rs\theta^2_T$	$rs\theta^2_T$	$s\sigma^2_{r(T)}$	$\sigma^2_{s(r(T))}$
$r(T)$	r	$s\sigma^2_{r(T)}$	~~$rs\theta^2_T$~~	$s\sigma^2_{r(T)}$	$\sigma^2_{s(r(T))}$
$s(r(T))$	s	$\sigma^2_{s(r(T))}$	~~$rs\theta^2_T$~~	~~$s\sigma^2_{r(T)}$~~	$\sigma^2_{s(r(T))}$

Źródło: opracowanie własne.

Ostateczną listę komponentów tworzących oczekiwany średni kwadrat dla każdego ze źródeł zaprezentowano w tabeli 10[16].

Tabela 10. Realizacja kroku III d dla schematu 2a. Oczekiwane średnie kwadraty dla schematu z efektem stałym T i efektem losowym r w strukturze zagnieżdżonej $r(T)$

Krok I	Krok II a	Krok II c	Krok III d		
Źródła	Poziomy	Komponenty	Oczekiwane średnie kwadraty $E(MS)$		
T	t	$rs\theta^2_T$	$rs\theta^2_T$	$+ s\sigma^2_{r(T)}$	$+ \sigma^2_{s(r(T))}$
$r(T)$	r	$s\sigma^2_{r(T)}$		$s\sigma^2_{r(T)}$	$+ \sigma^2_{s(r(T))}$
$s(r(T))$	s	$\sigma^2_{s(r(T))}$			$\sigma^2_{s(r(T))}$

Źródło: opracowanie własne.

Na początku tego podrozdziału wspomniano o tym, że wielu autorów zaproponowało różne metody mechanicznego generowania $E(MS)$.

[16] Gdzie (Sahai, Ageel, 2000: 352):

$$\theta^2_T = \frac{\sum(\alpha_i)^2}{(t-1)}$$

natomiast $\sigma^2_{r(T)}$ oraz $\sigma^2_{s(r(T))}$ to komponenty wariancyjne.

Różnice między tymi metodami odnotowywano głównie w przypisach powyższego fragmentu. Dodać można, że jeśli cel badacza byłby ograniczony jedynie do ustalenia odpowiedniego składnika błędu (mianownika ilorazu F), to postępowanie zaprezentowane wyżej można byłoby skrócić. Do realizacji tego celu wystarczy bowiem ustalenie zawartości indeksów dolnych poszczególnych komponentów, a wtedy możliwe byłoby pominięcie wszystkich czynności w obrębie kroku II. Zasady prowadzące do tak pojętego minimalistycznego rezultatu sformułowali Keppel i Wickens (2004) oraz Hopkins (1976). Dla porządku uzupełnijmy, że tym rezultatem w przypadku schematu 2b byłyby ustalenia zawarte w tabeli 11., a dla schematu 2a – w tabeli 12. Dla każdego efektu dodatkowo podano informację o odpowiadającym mu składniku błędu oraz postaci statystyki F.

Tabela 11. Źródła zmienności kształtujące średnie kwadraty dla poszczególnych efektów w schemacie 2b (modelu $T \times r$)

Źródło / efekt	MS	Źródła kształtujące MS	Składnik błędu	Iloraz F
T	MS_T	T, $r \times T$, $s(r \times T)$	$MS_{r\times T}$	$F = \frac{MS_T}{MS_{r\times T}}$
r	MS_r	r, $s(r \times T)$	$MS_{s(r\times T)}$	$F = \frac{MS_r}{MS_{s(r\times T)}}$
$r \times T$	$MS_{r\times T}$	$r \times T$, $s(r \times T)$	$MS_{s(r\times T)}$	$F = \frac{MS_{r\times T}}{MS_{s(r\times T)}}$
$s(r \times T)$	$MS_{s(r\times T)}$	$s(r \times T)$	–	–

Źródło: opracowanie własne.

Tabela 12. Źródła zmienności kształtujące średnie kwadraty dla poszczególnych efektów w schemacie 2a (modelu $r(T)$)

Źródło / efekt	MS	Źródła kształtujące MS	Składnik błędu	Iloraz F
T	MS_T	T, $r(T)$, $s(r(T))$	$MS_{r(T)}$	$F = \frac{MS_T}{MS_{r(T)}}$
$r(T)$	$MS_{r(T)}$	$r(T)$, $s(r(T))$	$MS_{s(r(T))}$	$F = \frac{MS_{r(T)}}{MS_{s(r(T))}}$
$s(r(T))$	$MS_{s(r(T))}$	$s(r(T))$	–	–

Źródło: opracowanie własne.

2.2.1. Źródła zmienności kształtujące wartości średnie dla czynnika stałego w schematach z replikacjami

Wygenerowanie $E(MS)$ dla czynnika T doprowadziło m.in. do ustalenia, jakie źródła zmienności kształtują MS_T w schemacie 2a, a jakie w schemacie 2b. Poniższy fragment ma na celu zobrazowanie tych ustaleń[17].

Oczywiste jest, że wielkość MS_T jest funkcją wielkości efektu T w populacji. To, że bezpośrednim pomiarom w eksperymencie objęta jest próbka, a nie populacja jednostek jest równie wiadome, podobnie jak to, że wyniki w próbie mogą być obciążone innymi błędami niż błąd losowania. Stąd zrozumiałe – choćby intuicyjnie – powinno być również to, że średnie $\bar{y}_i$ w próbie, a więc i MS_T, zależą także od wielkości błędu eksperymentalnego przypisywanego do jednostek s. Mniej oczywiste może być to, że dodatkowym elementem wpływu na wielkość MS_T może być $r(T)$ w schemacie 2a, oraz $T \times r$ w schemacie 2b (tabele 11. i 12.).

Żeby nieco ożywić rozważania, wróćmy do problemu skuteczności dwóch metod nauczania, który był wykorzystywany w rozdziale I. Dla ciągłości prowadzonego tutaj wywodu oznaczmy teraz czynnik „metoda nauczania" jako T. W mocy pozostają ustalenia, że jest to czynnik stały o dwóch poziomach, tutaj t_1 i t_2. W analizie uwzględnieni są również „nauczyciele" jako czynnik losowy r (będziemy opierać szacunki na podstawie próbki nauczycieli) oraz „uczniowie" jako czynnik losowy s.

Rysunek 5. ilustruje przypadek schematu 2a, w którym czynniki „nauczyciel" i „metoda nauczania" tworzą strukturę zagnieżdżoną $r(T)$. Wyniki na poziomie uczniów nie są pokazane – rysunek przedstawia dane na wyższym poziomie agregacji, tj. na poziomie nauczycieli. Dla każdego z 20 nauczycieli dysponujemy wynikiem, który jest średnią liczbą punktów zdobytych w teście przez uczniów należących do jego klasy. Na potrzeby obecnych rozważań zakładamy, że nauczyciele oznaczeni numerami od r_1 do r_{10} stanowią (pod)populację nauczających metodą t_1, a nauczyciele oznaczeni numerami od r_{11} do r_{20} stanowią (pod)populację nauczających metodą t_2.

Przedstawiony na rysunku 5. układ wyników ilustruje przypadek, w którym prawdziwa jest hipoteza zerowa, tj. metody nie różnią się skutecznością, $\mu_1 = \mu_2 = 6$, efekt T w populacji nie występuje (pokazuje to przedostatnia kolumna). Teraz wyobraźmy sobie, że z każdej (pod)populacji wylosowano po trzech nauczycieli. Na rysunku wyniki tych wylosowanych osób są obwiedzione ciągłą linią. W ostatniej kolumnie przedstawiono wartości średnie występujące w grupach t_1 i t_2, obliczone tym

[17] Inspirację do wywodu znaleziono w: Keppel, Wickens (2004: 534, 559); Maxwell, Delaney i Kelly (2018: 553–555); oraz Jackson, Brashers (1994a: 25–26).

razem na podstawie wyników nauczycieli, którzy dostali się do prób. I te wartości, $\bar{y}_1 = 6$ i $\bar{y}_2 = 7$, sugerują, że metoda t_2 jest bardziej skuteczna niż t_1. Gdyby do prób dostały się inne osoby, średnie przedstawiałyby się inaczej, a z każdym następnym losowaniem mogłyby sugerować inny wniosek na temat skuteczności porównywanych metod. Obserwujemy zatem sytuację, w której średnie $\bar{y}_1$ i $\bar{y}_2$ wykazują zmienność mimo, iż zmienność pochodząca od czynnika T wynosi zero. Zmienność średnich $\bar{y}_1$ i $\bar{y}_2$ pojawia się dlatego, że w obrębie grupy t_1, a także t_2, nauczyciele mają różne wyniki dotyczące skuteczności, innymi słowy z tego powodu, że zmienność pochodząca od czynnika $r(T)$ jest większa od zera (zakładamy również, że zmienność pochodząca od czynnika $s(r(T))$ jest większa od zera, ale tej zmienności akurat rysunek nie uwidacznia). Mówiąc jeszcze inaczej, odnotowywane w próbach średnie dla poszczególnych metod różnią się, ponieważ w obrębie poszczególnych metod występuje efekt nauczyciela (widoczny w dolnym wierszu).

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}	średnie w populacjach	średnie w próbach
t_1	4	5	7	8	8	4	5	7	6	6											6	6
t_2											2	5	7	10	6	4	8	6	3	9	6	7
średnie	4	5	7	8	8	4	5	7	6	6	2	5	7	10	6	4	8	6	3	9		

Rysunek 5. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 1.

Źródło: opracowanie własne.

Niezerowa zmienność pochodząca od $r(T)$ będzie wpływać na średnie grupowe $\bar{y}_i$ oczywiście także wtedy, gdy hipoteza zerowa jest fałszywa, a więc gdy efekt czynnika T występuje w populacji. Przykładowy układ wyników w takim przypadku ilustruje rysunek 6.

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}	średnie w populacjach	średnie w próbach
t_1	4	5	7	8	8	4	5	7	6	6											6	7
t_2											9	8	7	10	9	6	8	6	8	9	8	8
średnie	4	5	7	8	8	4	5	7	6	6	9	8	7	10	9	6	8	6	8	9		

Rysunek 6. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 2.

Źródło: opracowanie własne.

Dopiero przy braku efektu $r(T)$, czyli w sytuacji braku zróżnicowania wyników w populacji nauczających metodą t_1 oraz braku zróżnicowania wyników w populacji nauczających metodą t_2, zdanie się na próbki nauczycieli nie będzie niosło zagrożenia, że średnie w próbach będą inne od średnich w populacjach (zostawiając na boku kwestię zmienności pochodzącej od $s(r(T))$). Taką sytuację przedstawiają rysunki 7. i 8. Rysunek 7. zakłada dodatkowo wystąpienie efektu T w populacji, a rysunek 8. jego brak[18].

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}	średnie w populacjach	średnie w próbach
t_1	6	6	6	6	6	6	6	6	6	6											6	6
t_2											8	8	8	8	8	8	8	8	8	8	8	8
średnie	6	6	6	6	6	6	6	6	6	6	8	8	8	8	8	8	8	8	8	8		

Rysunek 7. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 3.

Źródło: opracowanie własne.

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	r_{11}	r_{12}	r_{13}	r_{14}	r_{15}	r_{16}	r_{17}	r_{18}	r_{19}	r_{20}	średnie w populacjach	średnie w próbach
t_1	6	6	6	6	6	6	6	6	6	6											6	6
t_2											6	6	6	6	6	6	6	6	6	6	6	6
średnie	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6		

Rysunek 8. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 4.

Źródło: opracowanie własne.

[18] Na marginesie warto dodać, że rysunki 5–8 można potraktować jako prototypowe w odniesieniu do jednoczynnikowego schematu międzygrupowego. Schemat ten ponownie zawierałby czynnik stały T (metoda nauczania), natomiast czynnik losowy $r(T)$ odpowiadałby teraz nie nauczycielom (zostają usunięci z modelu), tylko jednostkom badanym (tutaj uczniom) zagnieżdżonym w poszczególnych metodach nauczania. Wtedy wartości w środku tabeli należałoby potraktować nie jako średnie dla nauczycieli, tylko jako wyniki surowe jednostek badanych. Rozumowanie zaprezentowane w tekście dostarcza uzasadnienia, dlaczego statystyka F dla efektu T miałaby postać $F = MS_T/MS_{r(T)}$, co w tym przypadku należałoby rozumieć jako $F = MS_{metoda}/MS_{badane\ jednostki\ (metoda)}$.

Dopasujmy teraz powyższą sytuację problemową do schematu 2b, w którym czynniki „nauczyciel" i „metoda nauczania" tworzą strukturę krzyżową $r \times T$. Tak jak poprzednio, dysponujemy dziesięcioma średnimi wynikami dla metody t_1 oraz dziesięcioma średnimi dla metody t_2, ale z uwagi na to, że teraz każdy z nauczycieli uczy obiema metodami, obecnie nauczycieli jest łącznie dziesięciu, a nie dwudziestu, jak ostatnim razem. Na potrzeby dalszych rozważań zakładamy, że dziesięciu nauczycieli stanowi populację. W dalszym kroku będziemy losować z tej populacji próbki składające się z trzech nauczycieli.

Rysunek 9. przedstawia sytuację, w której efekt metody T nie występuje w populacji. Dane dodatkowo zaaranżowano w ten sposób, by nie występował efekt czynnika r, czyli aby nie występował efekt nauczyciela (średnie w dolnym wierszu są takie same). Zrobiono to w celu oczyszczenia obrazu sytuacji oraz zapobieżenia ewentualnej i niepożądanej sugestii, że efekt metody obserwowany na poziomie próby mógł być wywołany przez efekt nauczyciela. W odróżnieniu od poprzedniego schematu, w schemacie 2b to nie efekt r jest odpowiedzialny za to, że średnie dla metod w próbie fluktuują. Po wylosowaniu do próby nauczycieli r_1, r_2 i r_5 (zgodnie z widocznym na rysunku wyróżnieniem za pomocą ciągłej linii), średnie w próbie pokazują większą skuteczność metody t_2 w porównaniu z t_1. Inna próbka – powiedzmy r_6, r_7, r_{10} – sugerowałaby tę samą skuteczność obu metod. Wspomniana fluktuacja pojawia się w związku z wystąpieniem efektu interakcji $r \times T$, czyli z sytuacją, w której efekt metody jest inny dla każdego nauczyciela (i nie odzwierciedla zerowego efektu metody w populacji), np. r_1 i r_2 uzyskują wyższe wyniki, gdy nauczają metodą t_2, z kolei r_3 i r_4 uzyskują lepsze wyniki, gdy nauczają za pomocą metody t_1, itd. Tak długo jak poszczególni nauczyciele będą różnić się efektem metody, tak długo średnie $\bar{y}_1$ i $\bar{y}_2$ będą odzwierciedlać te wpływy (oprócz ewentualnych wpływów ze strony czynnika T oraz $s(r \times T)$, co jednak nie będzie już dalej demonstrowane[19]).

[19] Wprawdzie w książce tej nie omawiane są schematy wewnątrzgrupowe, ale na marginesie warto dodać, że rysunek 9. można potraktować jako prototypowy w odniesieniu do jednoczynnikowego schematu wewnątrzgrupowego. Schemat ten ponownie zawierałby czynnik stały T (metoda nauczania), natomiast czynnik losowy r odpowiadałby teraz nie nauczycielom (zostają usunięci z modelu), ale badanym osobom (tutaj uczniom). Wtedy wartości w środku tabeli należałoby potraktować nie jako średnie nauczycieli, tylko jako wyniki surowe osób badanych. Rozumowanie zaprezentowane w tekście dostarcza uzasadnienia, dlaczego statystyka F dla efektu T przyjęłaby postać $F = MS_T/MS_{r\times T}$, co w tym przypadku należałoby rozumieć jako $F = MS_{metoda}/MS_{badane\ jednostki \times metoda}$.

	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	średnie w populacji	średnie w próbie
t_1	4	5	7	8	6	2	10	4	8	6	6	5
t_2	8	7	5	4	6	10	2	8	4	6	6	7
średnie	6	6	6	6	6	6	6	6	6	6		

Rysunek 9. Wynik losowania trzech elementów z jednej populacji

Źródło: opracowanie własne.

2.3. Statystyka quasi-*F*

Przypomnijmy, że konstruując iloraz F, do mianownika należy szukać takiego MS, którego wartość oczekiwana składa się z tych samych komponentów, co wartość oczekiwana dla testowanego efektu, poza komponentem dla efektu. Przyjmuje się tu *implicite*, że uda się znaleźć dokładnie jeden średni kwadrat, który spełnia ten warunek. Niestety nie zawsze się to udaje – problemów pod tym względem często dostarczają schematy, w których czynnik losowy znajduje się w strukturze krzyżowej. Przykład takiego schematu, wraz z odpowiadającymi mu $E(MS)$, przedstawia tabela 13.

Tabela 13. Oczekiwane średnie kwadraty dla schematu $s \times r(T)$

Źródła	MS	$E(MS)$				
T	MS_T	$rs\theta_T^2$	$+s\sigma_{r(T)}^2$		$+r\sigma_{T\times s}^2$	$+\sigma_{s\times r(T)}^2$
$r(T)$	$MS_{r(T)}$		$s\sigma_{r(T)}^2$			$+\sigma_{s\times r(T)}^2$
s	MS_s			$tr\sigma_s^2$		$+\sigma_{s\times r(T)}^2$
$T \times s$	$MS_{T\times s}$				$r\sigma_{T\times s}^2$	$+\sigma_{s\times r(T)}^2$
$s \times r(T)$	$MS_{s\times r(T)}$					$\sigma_{s\times r(T)}^2$

Źródło: opracowanie własne.

Schemat ten[20] będzie interesował nas przez wzgląd na trudność poddania testowi czynnika stałego T. Zauważmy, że odpowiednim

[20] W nawiązaniu do Wprowadzenia i przedstawionego tam badania nad szybkością czytania słów – ten model, zdaniem Clarka (1973), powinni przyjąć badacze (T to czynnik stały „części mowy" o poziomach „czasowniki" i „rzeczowniki", $r(T)$ to czynnik losowy „słowa" zagnieżdżony w czynniku T, s to czynnik losowy „osoby badane"). Przykład innego badania opartego na tym schemacie Czytelnik znajdzie u Keppela i Wickensa (2004: 546).

składnikiem błędu jest w tym przypadku MS, którego $E(MS) = s\sigma^2_{r(T)} + r\sigma^2_{T\times s} + \sigma^2_{s\times r(T)}$. Niestety na liście nie ma MS, który spełniłby ten warunek. To, co można w tej sytuacji zrobić, to znaleźć takie wyrażenie złożone z kilku dostępnych średnich kwadratów, które pozwoli uzyskać pożądaną sumę komponentów $E(MS)$. Tutaj ten cel osiągniemy za pomocą:

$$E\left(MS_{r(T)} + MS_{T\times s} - MS_{s\times r(T)}\right) = s\sigma^2_{r(T)} + r\sigma^2_{T\times s} + \sigma^2_{s\times r(T)}$$

Statystyką testową będzie teraz quasi-F. Przyjmie ona następującą postać (Keppel, Wickens, 2004: 546):

$$F' = \frac{MS_T}{MS_{r(T)} + MS_{T\times s} - MS_{s\times r(T)}}$$

Zauważmy, że licznik będzie się różnił od mianownika jedynie komponentem dla efektu, gdy iloraz quasi-F skonstruujemy także w inny sposób (Clark, 1973):

$$F'' = \frac{MS_T + MS_{s\times r(T)}}{MS_{r(T)} + MS_{T\times s}}$$

Obydwie statystyki są poprawne zbudowane, niemniej sam fakt, że istnieją dwie, a zatem każda może przynieść inny wynik i inne rozstrzygnięcie wobec hipotezy zerowej, bywa powodem krytycznego nastawienia wobec stosowania quasi-F (Jackson, Brashers 1994a). Z drugiej strony nie jest kwestią zupełnej dowolności dla badacza, czy zastosuje F', czy F'', gdyż każda z tych statystyk ma swoje ograniczenia.

F' może przyjąć wartość ujemną, gdyż w mianowniku przewiduje się odejmowanie. Odpowiedzią na ten problem jest właśnie F'' (Winer, 1971: 377).

Natomiast problem dotyczący obu statystyk wiąże się z liczbą stopni swobody. Nie dotyczy on „zwykłej" statystyki F, ponieważ tam zarówno liczba swobody dla licznika df_1, jak i dla mianownika df_2 to liczby całkowite, dlatego ustosunkowując się do hipotezy zerowej, wielkość statystyki F można bezpośrednio odnieść do rozkładu F. Natomiast jeśli wielkość w liczniku (/mianowniku) jest wyrażeniem złożonym z kilku średnich kwadratów, to liczba stopni swobody dla licznika (/mianownika) będzie musiała być zaokrąglana do najbliższej liczby całkowitej. To znaczy, że rozkład F jest tylko przybliżonym (mniej lub bardziej) rozkładem dla quasi-F.

Sposób wyznaczania liczby stopni swobody podaję za Winerem (1971: 377–382), który z kolei relacjonuje metodę Satterthwaite'a (1946).

Przedstawmy ogólną postać F'' za pomocą:

$$F'' = \frac{u+v}{w+x}$$

gdzie u, v, w, x symbolizują odpowiednie średnie kwadraty. Liczba stopni swobody dla tych średnich kwadratów będzie oznaczona odpowiednio przez df_u, df_v, df_w, df_x.

Liczba stopni swobody df_1na licznika F'' będzie określona wzorem:

$$df_1 = \frac{(u+v)^2}{\frac{u^2}{df_u} + \frac{v^2}{df_v}}$$

A liczba stopni swobody dla mianownika F'' będzie określona wzorem:

$$df_2 = \frac{(w+x)^2}{\frac{w^2}{df_w} + \frac{x^2}{df_x}}$$

Teraz przedstawmy ogólną postać F' za pomocą:

$$F' = \frac{u}{w+x-v}$$

Wtedy liczba stopni swobody dla licznika F' będzie wynosić:

$$df_1 = df_u$$

A liczba stopni swobody dla mianownika F' będzie wynosić:

$$df_2 = \frac{(w+x-u)^2}{\frac{w^2}{df_w} + \frac{x^2}{df_x} + \frac{v^2}{df_v}}$$

Wzór na liczbę stopni swobody można uogólnić do postaci:

$$df = \frac{(\sum a_i\, MS_i)^2}{\sum_i \frac{(a_i MS_i)^2}{df_i}}$$

gdzie a_i może przyjąć wartość 1 albo -1. Obliczając df_2 dla mianownika ilorazu F', to $a_1 = 1$, $a_2 = 1$, $a_3 = -1$.

Clark (1973) opisuje jeszcze jedną statystykę, która mogłaby być w tej sytuacji zastosowana, tj. $Min\ F'$. Statystyka ta, zdaniem Smitha (2005), powinna być traktowana jako wariant drugiego wyboru (najlepszym będzie F' lub F''), jako że jest nieco zbyt konserwatywna (także Wickens, Keppel, 1983).

2.4. Założenia modeli

W modelu jednoczynnikowym z efektem stałym *T*:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

w którym α_i reprezentuje efekt stały i-tego poziomu czynnika T, przyjmowane są następujące założenia[21] (Glass, Hopkins, 1996: 402; Sahai, Ageel, 2000: 12):

- w każdej podpopulacji błędy losowe (reszty) ϵ_{ij} są niezależne, mają rozkład normalny o średniej równej 0 i wspólnej wariancji σ_ϵ^2 (tutaj σ_ϵ^2 odpowiada $\sigma_{s(T)}^2$). Te trzy kluczowe założenia: niezależności, normalności oraz jednorodności wariancji błędów losowych można zapisać w skrótowej postaci jako: $\epsilon_{ij} \sim NID(0, \sigma_\epsilon^2)$;
- na parametry nakłada się ograniczenia mówiące, że $\sum \alpha_i = 0$.

W modelu dwuczynnikowym z efektami stałymi *T* oraz *R* w strukturze krzyżowej $T \times R$:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

w którym α_i reprezentuje efekt stały i-tego poziomu czynnika T, β_j reprezentuje efekt stały czynnika R, a $(\alpha\beta)_{ij}$ reprezentuje efekt stały interakcji i-tego poziomu czynnika T z j-tym poziomem czynnika R, przyjmowane są następujące założenia (Sahai, Ageel, 2000: 180–181):

- $\epsilon_{ijk} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma_{s(T\times R)}^2$;
- w odniesieniu do trzech efektów stałych nakłada się ograniczenia mówiące, że $\sum \alpha_i = \sum \beta_j = \sum (\alpha\beta)_{ij} = 0$.

[21] W literaturze przedmiotu założenia formułowane są bądź w odniesieniu do rozkładu y_{ij} (np. Wiktorowicz, Grzelak, Grzeszkiewicz-Radulska, 2020: 105) bądź rozkładu ϵ_{ij}. Pamiętając, że ϵ_{ij} możemy przedstawić jako $\epsilon_{ij} = y_{ij} - \mu - \alpha_i$ oraz, że zarówno μ, jak i α_i to parametry, a więc wartości stałe, to ϵ_{ij} jest prostą transformacją y_{ij}. Z tego powodu założenia dotyczące rozkładów y_{ij} i ϵ_{ij} są równoważne (Glass, Hopkins, 1996: 402).

W modelu jednoczynnikowym z efektem losowym *t*:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

w którym α_i reprezentuje teraz nie stały, ale losowy efekt i-tego poziomu czynnika t, przyjmuje się, że (Glass, Hopkins, 1996: 537, 543; Sahai, Ageel, 2000: 12):

- $\epsilon_{ij} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma_{s(t)}^2$;
- α_i są niezależne i mają rozkład normalny o średniej równej 0 i wariancji σ_α^2 (tutaj σ_α^2 odpowiada σ_t^2), w skrócie $\alpha_i \sim NID(0, \sigma_\alpha^2)$;
- α_i oraz ϵ_{ij} są wzajemnie niezależne.

W modelu dwuczynnikowym z efektami losowymi *t* oraz *r* w strukturze krzyżowej $t \times r$:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

w którym α_i reprezentuje losowy efekt i-tego poziomu czynnika t, β_j reprezentuje losowy efekt czynnika r, natomiast $(\alpha\beta)_{ij}$ reprezentuje losowy efekt interakcji i-tego poziomu czynnika t z j-tym poziomem czynnika r, przyjmuje się następujące założenia (Sahai, Ageel, 2000: 180–181):

- $\epsilon_{ijk} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma_{s(t\times r)}^2$;
- $\alpha_i \sim NID(0, \sigma_\alpha^2)$, przy czym tutaj σ_α^2 odpowiada σ_t^2;
- $\beta_j \sim NID\left(0, \sigma_\beta^2\right)$, przy czym tutaj σ_β^2 odpowiada σ_r^2;
- $(\alpha\beta)_{ij} \sim NID\left(0, \sigma_{\alpha\beta}^2\right)$, przy czym tutaj $\sigma_{\alpha\beta}^2$ odpowiada $\sigma_{(t\times r)}^2$;
- α_i, β_j, $(\alpha\beta)_{ij}$ oraz ϵ_{ijk} są wzajemnie niezależne.

W modelu dwuczynnikowym z efektami losowymi *t* oraz *r* w strukturze zagnieżdżonej *r*(*t*):

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

w którym α_i reprezentuje losowy efekt i-tego poziomu czynnika t, $\beta_{j(i)}$ reprezentuje losowy efekt czynnika r zagnieżdżonego w czynniku t, przyjmuje się następujące założenia (Sahai, Ageel, 2000: 350, 371–373):

- $\epsilon_{k(ij)} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma_{s(r(t))}^2$;
- $\alpha_i \sim NID(0, \sigma_\alpha^2)$, przy czym tutaj σ_α^2 odpowiada σ_t^2;
- $\beta_{j(i)} \sim NID\left(0, \sigma_\beta^2\right)$, przy czym tutaj σ_β^2 odpowiada σ_r^2;
- α_i, $\beta_{j(i)}$, $\epsilon_{k(ij)}$ są wzajemnie niezależne.

W modelu dwuczynnikowym z efektami stałymi T oraz R w strukturze zagnieżdżonej $R(T)$:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

w którym α_i reprezentuje stały efekt i-tego poziomu czynnika T, $\beta_{j(i)}$ reprezentuje stały efekt czynnika R zagnieżdżonego w czynniku T, przyjmuje się następujące założenia (Sahai, Ageel, 2000: 350):

- $\epsilon_{k(ij)} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma^2_{s(R(T))}$;
- nakłada się ograniczenia mówiące, że $\sum \alpha_i = \sum \beta_{j(i)} = 0$.

W modelu dwuczynnikowym z efektem stałym T oraz efektem losowym r w strukturze zagnieżdżonej $r(T)$:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

w którym α_i reprezentuje stały efekt i-tego poziomu czynnika T, $\beta_{j(i)}$ reprezentuje losowy efekt czynnika r zagnieżdżonego w czynniku T, przyjmuje się następujące założenia (Sahai, Ageel, 2000: 350, 371-373):

- $\epsilon_{k(ij)} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma^2_{s(r(T))}$;
- $\beta_{j(i)} \sim NID\left(0, \sigma_\beta^2\right)$, przy czym tutaj σ_β^2 odpowiada $\sigma^2_{r(T)}$;
- $\beta_{j(i)}$ oraz $\epsilon_{k(ij)}$ są wzajemnie niezależne;
- nakłada się ograniczenie mówiące, że $\sum \alpha_i = 0$.

W modelu dwuczynnikowym z efektem stałym T oraz efektem losowym r w strukturze krzyżowej $T \times r$:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

w którym α_i reprezentuje efekt stały i-tego poziomu czynnika T, β_j reprezentuje losowy efekt czynnika r, natomiast $(\alpha\beta)_{ij}$ reprezentuje losowy efekt interakcji i-tego poziomu czynnika T z j-tym poziomem czynnika r, przyjmuje się następujące założenia (Stanisz, 2007: 592; Maxwell, Delaney, Kelly 2018: 557–558; Jackson, Brashers, 1994a: 33):

- $\epsilon_{ijk} \sim NID(0, \sigma_\epsilon^2)$, przy czym tutaj σ_ϵ^2 odpowiada $\sigma^2_{s(T\times r)}$;
- $\beta_j \sim NID\left(0, \sigma_\beta^2\right)$, przy czym tutaj σ_β^2 odpowiada σ_r^2;
- $\sum \alpha_i = 0$;
- $(\alpha\beta)_{ij} \sim N\left(0, \sigma^2_{\alpha\beta}\right)$, przy czym tutaj $\sigma^2_{\alpha\beta}$ odpowiada $\sigma^2_{T\times r}$, a także $\sum(\alpha\beta)_{ij} = 0$ dla wszystkich j. Założenie niezależności nie jest w tym przypadku spełnione;
- β_j, $(\alpha\beta)_{ij}$ oraz ϵ_{ijk} są wzajemnie niezależne.

Powyższe założenia dotyczące interakcji w modelu $T \times r$ są powszechnie przyjmowane w podręcznikach, ale trzeba wiedzieć, że są też inne sposoby ujęcia interakcji między czynnikiem losowym i stałym. Różnią się one, po pierwsze, obecnością ograniczenia $\sum(\alpha\beta)_{ij} = 0$ dla wszystkich j. Ograniczenie to uwzględnione jest w przyjętym tutaj zestawie założeń, a nawet więcej – w zasadach generowania $E(MS)$ przedstawionych w podrozdziale 2.2. Ograniczenie to głosi, że efekty interakcyjne sumują się do zera na wszystkich poziomach czynnika stałego. Ceną jego przyjęcia jest dopuszczenie do wystąpienia ujemnych korelacji między składnikami interakcji. Właśnie dlatego, by pozbyć się tej komplikacji niektórzy praktycy przyjmują model bez ograniczenia. Ale i to rozwiązanie ma swój koszt. Brak ograniczenia skutkuje tym, że $E(MS)$ dla czynnika r musiałby dodatkowo uwzględniać interakcję i przyjąć postać $E(MS_r) = ts\sigma_r^2 + \boldsymbol{s\sigma^2_{T\times r}} + \sigma^2_{s(T\times r)}$ (por. tabela 6.). To oznacza także konieczność zmiany składnika błędu w ilorazie F z $MS_{s(T\times r)}$ na $MS_{T\times r}$ (por. tabela 11.). Model bez ograniczenia nie pozwala więc sensownie ustosunkować się do hipotezy zerowej względem czynnika losowego. I właśnie ta utrata możliwości właściwej oceny efektu czynnika losowego r jest zasadniczym powodem, dla którego w podręcznikach akademickich tego modelu się nie wykorzystuje. Tymczasem Czytelnik powinien wiedzieć, że w domyślne opcje procedur analitycznych w SPSS, a także w SAS są zaimplementowane algorytmy odpowiadające modelowi bez ograniczenia (Maxwell, Delaney, Kelly, 2018: 605–606). To oznacza, że użytkownicy wymienionych pakietów, którzy są zainteresowani modelem z ograniczeniem powinni przeprowadzić analizy o samodzielnie napisane komendy w edytorze poleceń.

Druga różnica w przyjmowanych założeniach obejmuje jedynie modele z ograniczeniem. Dotyczy ona sposobu ujęcia wariancji efektów interakcyjnych. Tutaj przyjęto, że $(\alpha\beta)_{ij} \sim N\left(0, \sigma^2_{\alpha\beta}\right)$, natomiast niektórzy badacze przyjmują, że $(\alpha\beta)_{ij} \sim N\left(0, [(t-1)/t]\sigma^2_{\alpha\beta}\right)$ (np. Sahai, Ageel, 2000: 181). Ta różnica z kolei rzutuje na sposób oszacowania komponentów wariancyjnych (Ato, Vallejo, Palmer, 2013).

2.5. Miary wielkości efektu oparte na komponentach wariancyjnych

Najmniej skomplikowanym sposobem ustalenia wielkości efektu jest wyznaczenie R^2 (równoważnie: η^2). Wielkość ta obliczana jest jako stosunek sumy kwadratów odchyleń od średnich. Przykładowo wielkość efektu czynnika T obliczona byłaby jako (Maxwell, Delaney, Kelly, 2018: 564; Howell, 2007: 438):

$$R^2 = \frac{SS_T}{SS_{total}}$$

W przypadku modelu jednoczynnikowego wzór ten można ukonkretnić do postaci:

$$R^2 = \frac{SS_T}{SS_T + SS_{s(T)}}$$

a jego ogólną postać przedstawić jako:

$$R^2 = \frac{SS_{efekt\,T}}{SS_{efekt\,T} + SS_{błąd}}$$

W przypadku modelu dwuczynnikowego, przewidującego również obecność czynnika R, wielkość efektu dla czynnika T obliczona byłaby z wzoru:

$$R^2 = \frac{SS_T}{SS_{total}} = \frac{SS_T}{SS_T + SS_R + SS_{T\times R} + SS_{s(T\times R)}}$$

którego ogólna postać byłaby:

$$R^2 = \frac{SS_T}{SS_{total}} = \frac{SS_{efekt\,T}}{SS_{efekt\,T} + \sum SS_{pozostałe\,efekty} + SS_{błąd}}$$

W modelu dwu- i więcej czynnikowym możliwe byłoby też obliczenie cząstkowej wielkości efektu czynnika T:

$$R^2_{cząstkowy} = \frac{SS_{efekt\,T}}{SS_{efekt\,T} + SS_{błąd}} = \frac{SS_T}{SS_T + SS_{s(T\times R)}}$$

Proste od strony rachunkowej R^2 oraz $R^2_{cząstkowy}$ są dobrymi statystykami opisowymi, ale niestety jako oszacowania wielkości efektu w populacji są estymatorami obciążonymi. Na dodatek nie pozwalają różnicować wielkości efektów stałych oraz losowych (Howell, 2007: 438). Dzieje się tak, ponieważ wielkość SS dla czynnika będzie taka sama niezależnie od tego, czy ma on status stały czy losowy (np. $SS_T = SS_t$).

Zważywszy na te wady, w praktyce badawczej jako miary wielkości efektów stałych stosuje się ω^2 (omega kwadrat) oraz $\omega^2_{cząstkowy}$, a dla

efektów losowych stosuje się ich odpowiedniki, którymi są współczynnik korelacji wewnątrzklasowej ρ_I (ang. intraclass correlation) oraz cząstkowy współczynnik korelacji wewnątrzklasowej $\rho_{I\ cząstkowy}$ (Maxwell, Delaney, Kelly, 2018: 565; Kirk, 2005). Wprawdzie ω^2 oraz ρ_I (oraz ich cząstkowe odpowiedniki) również są estymatorami obciążonymi, niemniej w znacznie mniejszym stopniu niż R^2 (Kirk, 2005; Vaughan, Corballis, 1969: 206)[22].

Interpretację wartości ω^2 podaje R. Kirk (2005: 5). Według tej propozycji, opracowanej na podstawie skali Cohena (1988), wartości rzędu 0.010 uznaje się za świadczące o małej wielkości efektu. O średniej wielkości efektu świadczą wartości rzędu 0.059, a o dużej wielkości wartości rzędu 0.138. Wielkości te należy traktować orientacyjnie i brać pod uwagę, że w różnych dziedzinach nauki te wielkości mogą być inne.

R. Kirk (2005) w cytowanym wyżej encyklopedycznym źródle obszernie omawia zarówno ω^2, jak i ρ_I, ale nie podaje informacji, jak należy interpretować wartości drugiej z tych miar. Stosowana przez niektórych badaczy[23] praktyka, by w formułach na wielkość efektu posługiwać się tylko jednym symbolem miary, a konkretnie ω^2 (rozwiązanie to zakłada, że formuły uwzględniają dystynkcję między efektami losowymi i stałymi, a ujednolicenie dotyczy jedynie nazwy miary efektu), mogłaby sugerować, że wartości ρ_I należy interpretować tak samo, jak wartości ω^2. Sugestia taka nie będzie jednak trafna, gdyż – przy tym samym zestawie danych – zmiana statusu czynnika ze stałego T na losowy t będzie wiązać się ze wzrostem wartości miernika. Gdyby rozpatrywać model jednoczynnikowy, to wielkość efektu czynnika T wynosząca przykładowo 0.155, będzie odpowiadać wielkości 0.268 dla czynnika t. Dla modelu jednoczynnikowego relację między ω^2 a ρ_I opisują poniższe formuły (Kirk 2005):

$$\widehat{\omega}^2 = \frac{\hat{\rho}_I}{2 - \hat{\rho}_I}$$

$$\hat{\rho}_I = \frac{2\widehat{\omega}^2}{\widehat{\omega}^2 + 1}$$

Interpretację wartości pośrednich ρ_I w kategoriach „efekt słaby", „umiarkowany", „duży" spotyka się w literaturze rzadko, a autorzy podejmujący tę kwestię powołują się na bardzo ogólną zasadę, zgodnie

[22] Przegląd innych, szczegółowych ograniczeń omega kwadrat i korelacji wewnątrzklasowej Czytelnik znajdzie u R. Kirka (2005).
[23] Należy do nich m.in. Howell (2007: 438–439).

z którą progowymi wartościami byłyby odpowiednio 0.2, 0.5, 0.75 (Stanisz, 2007: 361–362; Hedges, Hedberg, 2007: 77) lub wskazują na sposób oceny odpowiadający współczynnikowi korelacji r Pearsona (Cohen, 1988: 83), czyli odpowiednio 0.1, 0.3, 0.5 (Field, 2005: 5; McGraw, Wong, 1996: 39). W rozważaniach tych uwzględnić jeszcze trzeba, że współczynnik korelacji wewnątrzklasowej jest wykorzystywany w bardzo różnych celach, także do oceny rzetelności pomiaru, a w takich przypadkach dopiero wartości bardzo wysokie, bliskie jedności, będą satysfakcjonujące. Jednocześnie przy innych zastosowaniach czy na innych polach wartości ρ_I rzędu 0.2 lub niższe mogą już być uznane za znaczące (Hedges, Hedberg, 2007: 77). Ostatecznie więc należy zgodzić się ze Staniszem (2007: 361), że „nie ma skali na to, jak duży musi być efekt, aby być znaczący". A zatem – kończąc wątek – ocena wielkości efektu, czy to ρ_I, czy ω^2, powinna odbywać się z uwzględnieniem wyników badań przeprowadzanych w danym obszarze i wartości tam raportowane powinny stać się punktem odniesienia dla oceny efektu.

Wracając do ogólnej charakterystyki omawianych mierników, należy dodać, że oszacowania zarówno ω^2 (Gamst, Meyers, Guarino, 2008), jak i ρ_I (Kenny, Judd, 1986) mogą przyjąć wartości ujemne. W takich przypadkach uznaje się, że wielkość objaśnionej zmienności wynosi zero (Maxwell, Delaney, Kelly, 2018: 130).

Oszacowania dla ω^2 oraz ρ_I pośrednio opierają się oczekiwanych średnich kwadratach $E(MS)$, a co za tym idzie wzory na wielkość efektu będą się różniły w zależności od modelu. To oznacza również, że badaczowi przydatna będzie umiejętność samodzielnego wyznaczania konkretnych formuł na wielkość efektu w oparciu o wzór ogólny. Umiejętność ta będzie szczególnie pożądana, gdy w grę wchodzić będzie model rozbudowany i złożony – wariantów takich modeli może być bardzo dużo i trudno, aby każdy z nich był omawiany w podręcznikach.

Ogólny wzór na wielkość efektu ma postać (Maxwell, Delaney, Kelly, 2018: 565, Howell, 2007: 438):

$$\frac{\hat{\sigma}^2_{efekt}}{\hat{\sigma}^2_{efekt} + \sum \hat{\sigma}^2_{pozostałe\ efekty} + \hat{\sigma}^2_{błąd}}$$

a gdy interesuje nas wielkość cząstkowa efektu, wtedy stosowana jest formuła:

$$\frac{\hat{\sigma}^2_{efekt}}{\hat{\sigma}^2_{efekt} + \hat{\sigma}^2_{błąd}}$$

Obecność $\hat{\sigma}^2$ w powyższych formułach oznacza, że zakładają one, iż każdy z efektów modelu jest losowy. W przypadku, w którym efekt miałby być stały, symbol $\hat{\sigma}^2$ należy zastąpić przez symbol $\widehat{\Theta}^2$.

Oba rodzaje estymatorów – $\hat{\sigma}^2$ oraz $\widehat{\Theta}^2$ – nazywa się komponentami wariancyjnymi (ang. *variance components*), choć tylko pierwszy zasługuje na to miano w pełni, jako że odnosi się do efektów losowych *per se*[24].

Dalszy wywód zostanie podzielony na części według podstawowych typów modeli analitycznych. W każdej części zostanie pokazane, jak należy szacować poszczególne komponenty wariancyjne, a także jaką postać mają wzory na wielkość poszczególnych efektów. Ponieważ podstawą wyznaczenia komponentów wariancyjnych jest znajomość $E(MS)$ dla modelu, to wywód jest uzupełniany informacją na ten temat. Wszystkie omawiane modele odnoszą się do schematów dla grup niezależnych oraz równolicznych.

Model jednoczynnikowy z efektem losowym *t*

Oczekiwane średnie kwadraty dla poszczególnych źródeł w tym modelu przedstawia tabela 14.

Tabela 14. Oczekiwane średnie kwadraty dla schematu z efektem losowym *t*

Źródła	Poziomy	MS	$E(MS)$	
t	t	MS_t	$s\sigma_t^2$	$+\ \sigma_{s(t)}^2$
$s(t)$	s	$MS_{s(t)}$		$\sigma_{s(t)}^2$

Źródło: opracowanie własne.

W modelu tym wielkości możliwe do oszacowania to $\sigma_{s(t)}^2$ oraz σ_t^2. W przypadku $\sigma_{s(t)}^2$ sprawa jest prosta, możemy go oszacować bezpośrednio za pomocą:

$$\hat{\sigma}_{s(t)}^2 = M_{s(t)}$$

Jeżeli chodzi o σ_t^2, to sprawa jest odrobinę trudniejsza, ponieważ nie występuje on w izolacji – MS_t dostarcza szacunku całości $s\sigma_t^2+\sigma_{s(t)}^2$, której σ_t^2 jest tylko częścią. Trudności tej można jednak szybko zara-

[24] Komponent wariancyjny $\hat{\sigma}^2$ jest definiowany jako oszacowanie wariancji rozkładu średnich w populacji poziomów czynnika losowego. Właściwa analiza komponentów wariancyjnych obejmuje więc z definicji jedynie efekty losowe (Glass, Hopkins, 1996: 549). Należy zaznaczyć, że analiza komponentów wariancyjnych jako taka nie jest w tej książce omawiana.

dzić. Wyizolowanie i oszacowanie σ_t^2 będzie możliwe dzięki prostemu przekształceniu:

$$\hat{\sigma}_t^2 = (MS_t - MS_{s(t)})/s$$

Badany efekt t jest losowy, a zatem jego wielkość będzie szacowana za pomocą ρ_I według wzoru:

$$\hat{\rho}_{I:t} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{s(t)}^2} = \frac{\frac{(MS_t - MS_{s(t)})}{s}}{\frac{(MS_t - MS_{s(t)})}{s} + MS_{s(t)}}$$

Po dalszych przekształceniach algebraicznych wzór ten przybiera znaną postać (Maxwell, Delaney, Kelly, 2018: 598):

$$\hat{\rho}_{I:t} = \frac{(MS_t - MS_{s(t)})}{(MS_t + (s-1)MS_{s(t)})}$$

której uogólnienie można znaleźć w wielu podręcznikach (np. Kirk, 2005: 537) jako:

$$\hat{\rho}_I = \frac{(MS_{efekt} - MS_{błąd})}{(MS_{efekt} + (s-1)MS_{błąd})}$$

Warto uzupełnić, że współczynnik korelacji wewnątrzklasowej może być wykorzystywany do oceny spełnienia założenia o niezależności pomiarów. Ocena taka może być przeprowadzona przede wszystkim w odniesieniu do źródła, które grupuje obserwacje i tworzy wspomniane już zagrożenie *braku niezależności z powodu grup* (Kenny, Judt, 1986; Jackson, Brashers, 1994a: 33). Nie jest tu więc mowa o zmiennej odpowiadającej głównej manipulacji eksperymentalnej, która grupuje obserwacje w sposób pożądany i celowy, ale o źródłach dodatkowo grupujących obserwacje w sposób poniekąd „ukryty", np. uczniowie są pogrupowani w klasy, respondenci tworzą grupy „przyporządkowane" ankieterom. Współczynnik korelacji wewnątrzklasowej pozwoli ustalić, w jakim stopniu obserwacje wewnątrz grup są do siebie podobne. Wielkości do powyższego wzoru uzyskamy przeprowadzając jednoczynnikową analizę wariancji, w której czynnikiem będzie zmienna grupująca (np. klasa, ankieter). Dysponując wielkością $\hat{\rho}_I$, w kolejnym kroku można ustalić stopień, w jakim obciążone są wyniki efektem zmiennej grupującej (Groves, 1989; Jabkowski, 2015: 86–95). Przypomnijmy też,

że remedium na to obciążenie jest włączenie zmiennej grupującej do zasadniczej analizy.

Przy okazji warto zauważyć, że interpretacja wielkości współczynnika korelacji wewnątrzklasowej może być wyznaczona zarówno pytaniem o procent wyjaśnionej wariancji (za jaki procent całkowitego zróżnicowania odpowiada zróżnicowanie między poziomami czynnika t?), (np. Sahai, Ageel, 2000: 45), jak i pytaniem o stopień podobieństwa elementów należących do tej samej grupy czynnika t (np. Stanisz, 2007: 583). Przykładowo wyobraźmy sobie badanie, w którym badanymi jednostkami są uczniowie (s), a badaną cechą jest ocena otrzymana przez ucznia (np. procent przyznanych punktów za napisany esej). Problem badawczy dotyczy wpływu egzaminatorów (oznaczonych jako czynnik t) na przyznaną liczbę punktów. Zakładamy sytuację, w której z populacji egzaminatorów wylosowano ich próbę, a także, że eseje przyporządkowano egzaminatorom na drodze losowej. Na jednego egzaminatora przypada kilka esejów i jeden esej jest oceniany przez jednego egzaminatora – to tworzy sytuację, w której czynnik „uczeń" jest zagnieżdżony w czynniku „egzaminator". Mamy więc w tym schemacie dwa źródła zmienności: t oraz $s(t)$. I teraz przyjmijmy, że efekt czynnika t okazał się istotny statystycznie oraz że

$$\hat{\rho}_{I:t} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{s(t)}^2} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_y^2} = \frac{89.1}{89.1 + 72.4} = \frac{89.1}{161.5} = 0.551 = 55.1\%$$

Interpretując te wyniki moglibyśmy powiedzieć, po pierwsze, że egzaminatorzy różnią się istotnie statystycznie pod względem średniej ocen (ocena istotnie zależy od tego, kto ocenia). Po drugie, wyrażając wielkość tej zależności, można byłoby powiedzieć, że za około 55% zróżnicowania w ocenach odpowiadają różnice między egzaminatorami w sposobach oceniania, i ekwiwalentnie, że korelacja pomiędzy ocenami uczniów ocenianych przez tego samego egzaminatora wynosi 0.551, a zatem oceny uczniów ocenianych przez tę samą osobę wykazują znaczne podobieństwo. Im wartość współczynnika korelacji wewnątrzklasowej byłaby bliższa 1, tym to podobieństwo byłoby większe, i tym samym większy byłby wpływ egzaminatora na ocenę.

Model dwuczynnikowy z efektami losowymi t oraz r w strukturze krzyżowej $t \times r$

Oczekiwane średnie kwadraty dla poszczególnych źródeł w tym modelu przedstawia tabela 15.

Tabela 15. Oczekiwane średnie kwadraty dla schematu z efektami losowymi $t \times r$

Źródła	Poziomy	MS	$E(MS)$			
t	t	MS_t	$rs\sigma_t^2$		$+s\sigma_{t\times r}^2$	$+\sigma_{s(t\times r)}^2$
r	r	MS_r		$ts\sigma_r^2$	$+s\sigma_{t\times r}^2$	$+\sigma_{s(t\times r)}^2$
$t \times r$		$MS_{t\times r}$			$s\sigma_{t\times r}^2$	$+\sigma_{s(t\times r)}^2$
$s(t \times r)$	s	$MS_{s(t\times r)}$				$\sigma_{s(t\times r)}^2$

Źródło: opracowanie własne.

Oszacowania komponentów wariancyjnych (Vaughan, Corballis, 1969: 206) to:

$$\hat{\sigma}_{s(t\times r)}^2 = MS_{s(t\times r)}$$

$$\hat{\sigma}_{t\times r}^2 = (MS_{t\times r} - MS_{s(t\times r)})/s$$

$$\hat{\sigma}_r^2 = (MS_r - MS_{t\times r})/ts$$

$$\sigma_t^2 = (MS_t - MS_{t\times r})/rs$$

Wielkość efektu losowego t będzie szacowana za pomocą wzoru (Maxwell, Delaney, Kelly, 2018: 565):

$$\hat{\rho}_{I:t} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_r^2 + \hat{\sigma}_{t\times r}^2 + \hat{\sigma}_{s(t\times r)}^2}$$

Korelację cząstkową wyznaczymy według formuły:

$$\hat{\rho}_{I\ cząstkowy:t} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{s(t\times r)}^2}$$

Analogicznie zostanie wyznaczone oszacowanie wielkości efektu czynnika r oraz efektu interakcji $t \times r$. Dla porządku podajmy:

$$\hat{\rho}_{I:r} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_t^2 + \hat{\sigma}_r^2 + \hat{\sigma}_{t\times r}^2 + \hat{\sigma}_{s(t\times r)}^2}$$

$$\hat{\rho}_{I\ cząstkowy:r} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_{s(t\times r)}^2}$$

$$\hat{\rho}_{I:t\times r} = \frac{\hat{\sigma}^2_{t\times r}}{\hat{\sigma}^2_t + \hat{\sigma}^2_r + \hat{\sigma}^2_{t\times r} + \hat{\sigma}^2_{s(t\times r)}}$$

$$\hat{\rho}_{I\ cząstkowy:t\times r} = \frac{\hat{\sigma}^2_{t\times r}}{\hat{\sigma}^2_{t\times r} + \hat{\sigma}^2_{s(t\times r)}}$$

Model jednoczynnikowy z efektem stałym *T*

Oczekiwane średnie kwadraty dla poszczególnych źródeł w tym modelu przedstawia tabela 1. Tym razem wielkościami, jakie mogą być oszacowane w oparciu o $E(MS)$ są θ^2_T oraz $\sigma^2_{s(T)}$. Ich estymatorami będą:

$$\hat{\sigma}^2_{s(T)} = MS_{s(T)}$$

$$\hat{\theta}^2_T = (MS_T - MS_{s(T)})/s$$

Przypomnijmy też, że z dwóch szacunków tylko pierwszy jest szacunkiem komponentu wariancyjnego.

Z uwagi na to, że czynnik *T* jest stały, wielkość jego efektu w populacji będzie teraz szacowana za pomocą ω^2 według wzoru[25]:

$$\widehat{\omega}^2_T = \frac{\widehat{\Theta}^2_T}{\widehat{\Theta}^2_T + \hat{\sigma}^2_{s(T)}}$$

Jak widać powyższy wzór[26] nie zawiera szacunku $\hat{\theta}^2_T$, którym dysponujemy, natomiast zawiera szacunek komponentu wariancyjnego $\widehat{\Theta}^2_T$. Relację między tymi wielkościami określa formuła (Vaughan, Corballis, 1969: 206; Jackson, Brashers, 1994a: 39):

$$\Theta^2_T = \theta^2_T \frac{t-1}{t}$$

[25] Jest on równoważny do wzoru, który można często spotkać w podręcznikach (Szymczak, 2018: 321; Keppel, Wickens, 2004: 164):

$$\widehat{\omega}^2_T = \frac{SS_T - (t-1)MS_{s(T)}}{SS_{total} + MS_{s(T)}}$$

[26] Wzór ten można również przedstawić w postaci (Maxwell, Delaney, Kelly, 2018: 566):

$$\widehat{\omega}^2_T = \frac{\sum \hat{\alpha}^2_i /t}{\sum \hat{\alpha}^2_i /t + \hat{\sigma}^2_{s(T)}}$$

Przypomnijmy jednocześnie (patrz przypis 6.), że $\theta^2_T = \sum \alpha^2_i/(t-1)$, dlatego jego oszacowanie nie może być bezpośrednio użyte we wzorze.

Oszacowanie komponentu wariancyjnego potrzebne do wzoru na wielkość efektu uzyskamy zatem w następujący sposób:

$$\widehat{\Theta}_T^2 = \frac{(t-1)\,(MS_T - MS_{s(T)})}{st}$$

A zatem wielkość efektu stałego T będzie ostatecznie szacowana za pomocą wzoru:

$$\widehat{\omega}_T^2 = \frac{\frac{(t-1)\,(MS_T - MS_{s(T)})}{st}}{\frac{(t-1)\,(MS_T - MS_{s(T)})}{st} + MS_{s(T)}}$$

Model dwuczynnikowy z efektami stałymi T oraz R w strukturze krzyżowej $T \times R$

Oczekiwane średnie kwadraty dla poszczególnych źródeł w tym modelu są przedstawione w tabeli 2. Na podstawie $E(MS)$ można dokonać następujących oszacowań, z których tylko pierwszy jest komponentem wariancyjnym:

$$\hat{\sigma}_{s(T\times R)}^2 = MS_{s(T\times R)}$$

$$\hat{\theta}_T^2 = \frac{MS_T - MS_{s(T\times R)}}{rs}$$

$$\hat{\theta}_R^2 = \frac{MS_R - MS_{s(T\times R)}}{ts}$$

$$\hat{\theta}_{T\times R}^2 = \frac{MS_{T\times R} - MS_{s(T\times R)}}{s}$$

Skoro mamy do czynienia z trzema efektami stałymi, do wzorów na wielkość efektów potrzebne będą także odpowiednie oszacowania dla Θ^2. Biorąc pod uwagę, że:

$$\Theta_T^2 = \theta_T^2 \frac{t-1}{t}$$

$$\Theta_R^2 = \theta_R^2 \frac{r-1}{r}$$

$$\Theta_{T\times R}^2 = \theta_{T\times R}^2 \frac{(t-1)(r-1)}{tr}$$

pozostałe komponenty wariancyjne uzyskamy za pomocą następujących oszacowań (Vaughan, Corballis, 1969: 206):

$$\widehat{\Theta}_T^2 = \frac{(t-1)(MS_T - MS_{s(T\times R)})}{trs}$$

$$\widehat{\Theta}_R^2 = \frac{(r-1)(MS_R - MS_{s(T\times R)})}{trs}$$

$$\widehat{\Theta}_{T\times R}^2 = \frac{(t-1)(r-1)(MS_{T\times R} - MS_{s(T\times R)})}{trs}$$

Wielkość efektu dla czynnika T będzie oszacowana za pomocą wzoru:

$$\widehat{\omega}_T^2 = \frac{\widehat{\Theta}_T^2}{\widehat{\Theta}_T^2 + \widehat{\Theta}_R^2 + \widehat{\Theta}_{T\times R}^2 + \hat{\sigma}_{s(T\times R)}^2}$$

a cząstkowa wielkość efektu T według wzoru:

$$\widehat{\omega}_{cząstkowy:\,T}^2 = \frac{\widehat{\Theta}_T^2}{\widehat{\Theta}_T^2 + \hat{\sigma}_{s(T\times R)}^2}$$

Na analogicznej zasadzie (za pomocą ω^2 lub $\omega_{cząstkowy}^2$) będą wyznaczane wielkości efektów dla czynnika R oraz interakcji $T \times R$.

Model dwuczynnikowy z efektem stałym T oraz efektem losowym r w strukturze krzyżowej $T \times r$

Powyższy model odpowiada schematowi 2b. Oczekiwane średnie kwadraty w tym modelu przedstawia tabela 6. Komponenty wariancyjne, jakie można na tej podstawie wyznaczyć dotyczą tylko efektów losowych. Będą to (Maxwell, Delaney, Kelly, 2018: 566; Vaughan, Corballis, 1969: 206; Sahai, Ageel, 2000: 206)[27]:

$$\hat{\sigma}_{s(T\times r)}^2 = MS_{s(T\times r)}$$

[27] W przeciwieństwie autorów wymienionych w odwołaniu, Howell (2007: 439–440) wyznacza komponent wariancyjny dla interakcji $T \times r$ za pomocą formuły:

$$\widehat{\Theta}_{T\times r}^2 = \frac{(t-1)(MS_{T\times r} - MS_{s(T\times r)})}{ts}$$

Można przypuszczać, że wiąże się to z przyjęciem przez Howella założenia, że $(\alpha\beta)_{ij} \sim N\left(0, [(t-1)/t]\sigma_{\alpha\beta}^2\right)$ (zob. podrozdział 2.4).

$$\hat{\sigma}_r^2 = \frac{MS_r - MS_{s(T\times r)}}{ts}$$

$$\hat{\sigma}_{T\times r}^2 = \frac{MS_{T\times r} - MS_{s(T\times r)}}{s}$$

Oszacowanie komponentu dla efektu stałego T będzie wyznaczone za pomocą:

$$\hat{\theta}_T^2 = \frac{MS_T - MS_{T\times r}}{rs}$$

a odpowiadający mu komponent wariancyjny będzie oszacowany w sposób:

$$\widehat{\Theta}_T^2 = \frac{(t-1)(MS_T - MS_{T\times r})}{trs}$$

Do oszacowania wielkości efektu stałego T w populacji posłuży wzór (Maxwell, Delaney, Kelly, 2018: 566):

$$\widehat{\omega}_T^2 = \frac{\widehat{\Theta}_T^2}{\widehat{\Theta}_T^2 + \widehat{\sigma}_r^2 + \widehat{\sigma}_{T\times r}^2 + \hat{\sigma}_{s(T\times r)}^2}$$

W celu oszacowania $\omega^2_{cząstkowy:\,T}$, należy w mianowniku pominąć $\widehat{\sigma}_r^2$ oraz $\widehat{\sigma}_{T\times r}^2$.

Natomiast wielkości efektów losowych, czyli r oraz $T \times r$ będą wyznaczane za pomocą wzorów na korelację wewnątrzklasową (Maxwell, Delaney, Kelly, 2018: 566):

$$\hat{\rho}_{I:r} = \frac{\hat{\sigma}_r^2}{\widehat{\Theta}_T^2 + \hat{\sigma}_r^2 + \hat{\sigma}_{T\times r}^2 + \hat{\sigma}_{s(\times r)}^2}$$

$$\hat{\rho}_{I\ cząstkowy:r} = \frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + \hat{\sigma}_{s(T\times r)}^2}$$

$$\hat{\rho}_{I:T\times r} = \frac{\hat{\sigma}_{T\times r}^2}{\widehat{\Theta}_T^2 + \hat{\sigma}_r^2 + \hat{\sigma}_{T\times r}^2 + \hat{\sigma}_{s(T\times r)}^2}$$

$$\hat{\rho}_{I\ cząstkowy:T\times r} = \frac{\hat{\sigma}_{T\times r}^2}{\hat{\sigma}_{T\times r}^2 + \hat{\sigma}_{s(T\times r)}^2}$$

Model dwuczynnikowy z efektem stałym T oraz efektem losowym r w strukturze zagnieżdżonej $r(T)$

Ten model odpowiada schematowi 2a – oczekiwane średnie kwadraty dla źródeł zmienności w tym schemacie przedstawia tabela 10. Na jej podstawie można oszacować następujące komponenty wariancyjne (Maxwell, Delaney, Kelly, 2018: 588–589, 600):

$$\hat{\sigma}^2_{s(r(T))} = MS_{s(r(T))}$$

$$\hat{\sigma}^2_{r(T)} = (MS_{r(T)} - MS_{s(r(T))})/s$$

a także komponent dla jednego efektu stałego:

$$\hat{\theta}^2_T = (MS_T - MS_{r(T)})/rs$$

Poniższe wyrażenie umożliwi przekształcenie tego komponentu w komponent wariancyjny:

$$\Theta^2_T = \theta^2_T \frac{t-1}{t}$$

który będzie oszacowany w sposób:

$$\widehat{\Theta}^2_T = \frac{(t-1)(MS_T - MS_{r(T)})}{trs}$$

Do oszacowania wielkości efektu stałego T w populacji posłuży wzór:

$$\widehat{\omega}^2_T = \frac{\widehat{\Theta}^2_T}{\widehat{\Theta}^2_T + \widehat{\sigma}^2_{r(T)} + \hat{\sigma}^2_{s(r(T))}}$$

W celu uzyskania oszacowania dla $\omega^2_{cząstkowy:\,T}$, należy w mianowniku pominąć $\widehat{\sigma}^2_{r(T)}$.

W celu oszacowania wielkości efektu losowego $r(T)$ należy posłużyć się formułą:

$$\hat{\rho}_{I:r(T)} = \frac{\hat{\sigma}^2_{r(T)}}{\widehat{\Theta}^2_T + \hat{\sigma}^2_{r(T)} + \hat{\sigma}^2_{s(T\times r)}}$$

Wyznaczając $\hat{\rho}_{I\ cząstkowy:r(T)}$, w mianowniku należy pominąć $\widehat{\Theta}^2_T$.

Model dwuczynnikowy z efektami losowymi t i r w strukturze zagnieżdżonej $r(t)$

Tabela 16. Oczekiwane średnie kwadraty dla schematu z efektami losowymi $r(t)$

Źródła	Poziomy	MS	$E(MS)$		
t	t	MS_t	$rs\sigma_t^2$	$+s\sigma_{r(t)}^2$	$+\sigma_{s(r(t))}^2$
$r(t)$	r	$MS_{r(t)}$		$s\sigma_{r(t)}^2$	$+\sigma_{s(r(t))}^2$
$s(r(t))$	s	$MS_{s(r(t))}$			$\sigma_{s(r(t))}^2$

Źródło: opracowanie własne.

Na podstawie oczekiwanych średnich kwadratów (tabela 16.) można dokonać oszacowań dla „docelowych" komponentów wariancyjnych (Maxwell, Delaney, Kelly, 2018: 589):

$$\hat{\sigma}_{s(r(t))}^2 = MS_{s(r(t))}$$

$$\hat{\sigma}_{r(t)}^2 = (MS_{r(t)} - MS_{s(r(t))})/s$$

$$\hat{\sigma}_t^2 = (MS_t - MS_{r(t)})/rs$$

Wielkość efektu losowego t będzie szacowana za pomocą wzoru:

$$\hat{\rho}_{I:t} = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{r(t)}^2 + \hat{\sigma}_{s(r(t))}^2}$$

Korelację cząstkową dla tego efektu wyznaczymy z pominięciem $\hat{\sigma}_{r(t)}^2$ w mianowniku.

Wielkość efektu losowego $r(t)$ będzie szacowana za pomocą:

$$\hat{\rho}_{I:r(t)} = \frac{\hat{\sigma}_{r(t)}^2}{\hat{\sigma}_t^2 + \hat{\sigma}_{r(t)}^2 + \hat{\sigma}_{s(r(t))}^2}$$

Natomiast korelację cząstkową dla tego efektu wyznaczymy z pominięciem $\hat{\sigma}_t^2$ w mianowniku.

Model dwuczynnikowy z efektami stałymi T i R w strukturze zagnieżdżonej $R(T)$

Dla porządku weźmy jeszcze pod uwagę model dwuczynnikowy z efektami stałymi T i R w strukturze zagnieżdżonej, choć rzadko w praktyce badawczej mamy do czynienia z sytuacją, w której czynnik zagnieżdżony jest stały. Oczekiwane średnie kwadraty w tym modelu przedstawia tabela 17[28].

Tabela 17. Oczekiwane średnie kwadraty dla schematu z efektami stałymi w strukturze zagnieżdżonej $R(T)$

Źródła	Poziomy	MS	$E(MS)$		
T	t	MS_T	$rs\theta_T^2$		$+\sigma^2_{s(R(T))}$
$R(T)$	r	$MS_{R(T)}$		$s\theta^2_{R(T)}$	$+\sigma^2_{s(R(T))}$
$s(R(T))$	s	$MS_{s(R(T))}$			$\sigma^2_{s(R(T))}$

Źródło: opracowanie własne.

Oszacowaniami, jakich można dokonać na podstawie oczekiwanych średnich kwadratów (tabela 17.) będą:

$$\sigma^2_{s(R(T))} = MS_{s(R(T))}$$

$$\hat{\theta}_T^2 = (MS_T - MS_{s(R(T))})/rs$$

$$\hat{\theta}^2_{R(T)} = (MS_{R(T)} - MS_{s(R(T))})/s$$

Dwa ostatnie oszacowania dotyczą efektów stałych, a więc wymagają przekształcenia do komponentów wariancyjnych. Biorąc pod uwagę, że:

$$\Theta_T^2 = \theta_T^2 \frac{t-1}{t}$$

[28] Gdzie (Sahai, Ageel 2000: 352, Maxwell, Delaney, Kelly, 2018: 585):

$$\theta_T^2 = \frac{\sum(\alpha_i)^2}{(t-1)}$$

$$\theta^2_{R(T)} = \frac{\sum\sum\beta^2_{j(i)}}{t \times (r-1)}$$

$$\Theta^2_{R(T)} = \theta^2_T \frac{r-1}{r}$$

komponenty wariancyjne będą miały postać (Maxwell, Delaney, Kelly, 2018: 589):

$$\widehat{\Theta}^2_T = \frac{(t-1)(MS_T - MS_{s(R(T))})}{trs}$$

$$\widehat{\Theta}^2_{R(T)} = \frac{(r-1)(MS_{R(T)} - MS_{s(R(T))})}{rs}$$

Wielkości dla efektów zostaną oszacowanie za pomocą ω^2:

$$\widehat{\omega}^2_T = \frac{\widehat{\Theta}^2_T}{\widehat{\Theta}^2_T + \widehat{\Theta}^2_{R(T)} + \hat{\sigma}^2_{s(R(T))}}$$

Aby oszacować $\omega^2_{cząstkowy:\,T}$, należy w mianowniku pominąć $\widehat{\Theta}^2_{R(T)}$.

$$\widehat{\omega}^2_{R(T)} = \frac{\widehat{\Theta}^2_{R(T)}}{\widehat{\Theta}^2_T + \widehat{\Theta}^2_{R(T)} + \hat{\sigma}^2_{s(R(T))}}$$

Wyznaczając $\widehat{\omega}^2_{cząstkowy:\,R(T)}$, w mianowniku należy pominąć $\widehat{\Theta}^2_T$.

W podrozdziale tym przedstawiono miary wielkości efektu oparte na komponentach wariancyjnych dla wszystkich typów modeli jedno- i dwuczynnikowych. Można wyrazić nadzieję, że wywód ten dał wystarczające podstawy do samodzielnego opracowania komponentów wariancyjnych w modelach trzy-[29] i więcej czynnikowych.

Dodajmy jeszcze, że omega kwadrat oraz korelacja wewnątrzklasowa nie muszą być miarami jedynie omnibusowymi – możliwe jest dokonanie oszacowań wielkości efektów dla kontrastów[30].

2.6. Moc testów

Przypomnijmy, że zagadnienie mocy testów podjęto już w paragrafach 1.2.3.3 oraz 1.2.3.4 przy okazji omawiania planu 2b (modelu z efektem stałym T oraz efektem losowym r w strukturze krzyżowej) oraz planu

[29] Komponenty wariancyjne dla modeli trzyczynnikowych z efektami stałymi, losowymi i mieszanymi w strukturze krzyżowej przedstawiają Vaughan i Corballis (1969: 208).
[30] Zainteresowanym Czytelnikom można polecić lekturę Keppela i Wickensa (2004).

2a (modelu z efektem stałym T oraz efektem losowym r w strukturze zagnieżdżonej). Przedstawiono tam m.in. czynniki, które w powyższych modelach wpływają na moc testu efektu T, a więc zwiększają lub zmniejszają możliwość wykrycia istniejącego efektu T. W obecnym fragmencie będzie podjęty inny aspekt zagadnienia mocy testów. Mianowicie zostaną przedstawione procedury umożliwiające ustalenie mocy testów dla poszczególnych efektów w różnych schematach.

Wydaje się, że najbardziej uniwersalny zestaw procedur analitycznych do wyznaczania mocy testów opracował Koele (1982). Propozycja ta powstała w oparciu o zasady sformułowane przez Scheffégo (1959). Obejmuje ona testy dla efektów stałych oraz losowych w schematach czynnikowych dla grup niezależnych, a jak przekonują Jackson i Brashers (1994a) może być też zastosowana do schematów, w których czynniki pozostają w strukturze zagnieżdżonej. Jest to jednocześnie propozycja, która daje się zastosować przy pomocy oprogramowania, w tym najbardziej dostępnego, jakim jest MS Excel i jego darmowe dodatki. Propozycja ta wymaga jednak umiejętności wyznaczania oczekiwanych średnich kwadratów (oraz komponentów wariancyjnych w przypadku ewentualnej potrzeby obliczenia wielkości mocy na podstawie wyników próby).

Moc testu dla efektu stałego[31] to prawdopodobieństwo, że wartość statystyki F będzie większa niż wartość krytyczna F (F_{kryt}), które jest określane przy danej liczbie stopni swobody dla licznika ilorazu F (df_1), liczbie swobody dla mianownika (df_2) oraz przy danej wartości niecentralnego parametru lambda (λ) (Koele, 1982):

$$1 - \beta = P(F > F_{kryt} \backslash df_1, df_2, \lambda)$$

Przypomnijmy, że F_{kryt} jest ustalana przy założeniu, że hipoteza zerowa jest prawdziwa. Jest wyznaczana[32] przy przyjętym przez badacza poziomie istotności ($alfa$) oraz przy danej liczbie stopni swobody dla licznika ilorazu F (df_1) oraz danej liczbie swobody dla mianownika (df_2).

Za Jackson i Brashers (1994a: 39) uogólniony wzór na λ można przedstawić jako:

[31] Moc testu dla efektu stałego może być wyznaczona przy pomocy funkcji NF_DIST dostępnej w ramach nieodpłatnego pakietu https://www.real-statistics.com/ będącego Dodatkiem do MS Excel. Moc może być obliczona według następującej formuły:
=1-NF_DIST(F kryt;df1;df2;lambda;skumulowany niecentralny rozkład F=1)

[32] Wartość F_{kryt} można obliczyć przy pomocy funkcji MS Excel:
=ROZKŁ.F.ODWR.PS(przyjęta wartość *alfa*, np. 0.05;df1;df2)

$$\lambda = n\frac{\Theta^2_{efekt}}{E(MS)_{błąd\ dla\ efektu}}$$

gdzie – przypomnijmy – n oznacza liczebność próby, Θ^2_{efekt} jest komponentem wariancyjnym dla badanego efektu (a ściślej odpowiednikiem komponentu wariancyjnego, gdyż chodzi o efekt stały)[33], a $E(MS)_{błąd\ dla\ efektu}$ to wartość oczekiwana średniego kwadratu dla błędu.

Istotne jest, że lambda będzie mogła być wyrażana w kategoriach $\boldsymbol{f^2}$, czyli **miary wielkości efektu** wprowadzonej przez Cohena (1988) i stosowanej przy ustalaniu mocy testu. Miara ta jest standaryzowana (wielkość efektu jest odnoszona do wariancji wewnątrzgrupowej). Wykorzystując dotychczas wprowadzone wielkości i oznaczenia, uogólnioną postać wzoru na f^2 dla efektu stałego w schematach dla grup niezależnych można przedstawić jako:

$$f^2_{efekt\ stały} = \frac{\Theta^2_{efekt}}{\sigma^2_{wewnątrzgrupowa}}$$

Powyższy wzór będzie wykorzystywany w dalszej części do wyznaczania lambdy dla poszczególnych efektów stałych w kolejnych schematach.

Natomiast do ustalenia mocy testu efektu losowego[34] wykorzystuje się κ razy centralny rozkład F (Koele, 1982):

$$1 - \beta = P\left(\kappa F > F_{kryt} \backslash df_1, df_2\right) = P\left(F > \frac{F_{kryt}}{\kappa} \backslash df_1, df_2\right)$$

Wielkość kappa (κ) wyznaczana jest zgodnie z formułą (Koelle 1982: 514; Jackson, Brashers 1994a: 39):

$$\kappa = \frac{E(MS)_{efekt}}{E(MS)_{błąd\ dla\ efektu}}$$

[33] Przekształcenia między komponentami wariancyjnymi Θ^2 a wariancjami θ^2 dla poszczególnych efektów stałych w kolejnych schematach Czytelnik znajdzie w podrozdziale 2.5.

[34] Moc testu dla efektu losowego może być obliczona przy pomocy funkcji ROZKŁ.F w MS Excel według następującej formuły
=1-ROZKŁ.F(F kryt /kappa;df1;df2;skumulowany=PRAWDA)

Podobnie jak lambda, także kappa będzie mogła być wyrażana w kategoriach f^2. Uogólniona postać wzoru na f^2 dla efektu losowego wygląda następująco:

$$f^2_{efekt\ losowy} = \frac{\sigma^2_{efekt}}{\sigma^2_{wewnątrzgrupowa}}$$

Powyższy wzór będzie wykorzystywany w dalszej części do wyznaczania kappy dla poszczególnych efektów losowych w kolejnych schematach.

Zgodnie z propozycją Cohena, wartości f^2 rzędu 0.01 ($f = 0.1$) świadczą o małej wielkości efektu, f^2 rzędu 0.0625 ($f = 0.25$) świadczą o średniej wielkości efektu, a wartości f^2 rzędu 0.16 ($f = 0.4$) świadczą o dużej wielkości efektu (1988: 285–287). Te trzy „progowe" wielkości zostały wykorzystane w tabelach 27–32 aneksu[35].

Z uwagi na to, że moc testu – co do zasady – powinna być wyznaczana *apriori*, to znaczy przed rozpoczęciem badania, w oparciu o zakładane wartości parametrów, a nie ich szacunki, to w formułach dalej przedstawionych eksponowane są parametry[36].

W przypadku ewentualnej potrzeby wyznaczenia mocy w oparciu o szacunki przydatna byłaby informacja o relacjach między $\hat{f}^2$ a miarami wielkości efektu przedstawionymi w poprzednim podrozdziale. Dla modeli jednoczynnikowych zachodzą relacje (Kirk, 1996):

$$\hat{f}^2_{efekt\ stały} = \frac{\widehat{\omega}^2}{1 - \widehat{\omega}^2}$$

$$\widehat{\omega}^2 = \frac{\hat{f}^2_{efekt\ stały}}{1 + \hat{f}^2_{efekt\ stały}}$$

$$\hat{f}^2_{efekt\ losowy} = \frac{\hat{\rho}_I}{1 - \hat{\rho}_I}$$

$$\hat{\rho}_I = \frac{\hat{f}^2_{efekt\ losowy}}{1 + \hat{f}^2_{efekt\ losowy}}$$

[35] W nawiązaniu do podrozdziału 2.5, w sytuacji, w której badacz uzna, że skale wartości pozwalające ocenić efekt w kategoriach słaby – umiarkowany – silny powinny być inne dla $\widehat{\omega}^2$ i $\hat{\rho}_I$, wtedy powinny także powstać dwie inne skale oceny dla $\hat{f}^2_{efekt\ stały}$ i $\hat{f}^2_{efekt\ losowy}$.

[36] Z dyskusją na temat kontrowersyjnej praktyki wyznaczania tzw. obserwowanej mocy (mocy „post hoc") Czytelnik może zapoznać się w pracy: Maxwell, Delaney, Kelly, 2018: 149–152.

Model jednoczynnikowy z efektem losowym *t*

W tym modelu kappa dla t powinna być wyznaczona za pomocą formuły:

$$\kappa_t = \frac{E(MS)_t}{E(MS)_{s(t)}} = \frac{s\sigma_t^2 + \sigma_{s(t)}^2}{\sigma_{s(t)}^2} = 1 + s\frac{\sigma_t^2}{\sigma_{s(t)}^2} = 1 + sf_t^2$$

W przypadku ewentualnej potrzeby oszacowania f_t^2 na podstawie wyników otrzymanych w próbie, oszacowanie takie uzyskamy dzięki komponentom wariancyjnym (Maxwell, Delaney, Kelly, 2018: 571), przedstawionym w poprzednim podrozdziale[37].

Model dwuczynnikowy z efektami losowymi *t* oraz *r* w strukturze krzyżowej $t \times r$

W tym modelu kappa dla efektu t zostanie wyznaczona za pomocą formuły:

$$\kappa_t = \frac{E(MS)_t}{E(MS)_{t\times r}} = \frac{rs\sigma_t^2 + s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}{s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}$$

Aby móc κ_t wyrazić w kategoriach f^2, należy licznik i mianownik podzielić przez wariancję błędu przypisanego do czynnika „osoby badane", czyli przez wariancję wewnątrzgrupową:

$$\kappa_t = \frac{\frac{rs\sigma_t^2 + s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}{\sigma_{s(t\times r)}^2}}{\frac{s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}{\sigma_{s(t\times r)}^2}} = \frac{1 + \frac{rs\sigma_t^2}{\sigma_{s(t\times r)}^2} + \frac{s\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}}{1 + \frac{s\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}} = \frac{1 + \frac{rs\sigma_t^2}{\sigma_{s(t\times r)}^2}}{1 + \frac{s\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}} +$$

$$+ \frac{\frac{s\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}}{1 + \frac{s\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}} = 1 + \frac{rs\frac{\sigma_t^2}{\sigma_{s(t\times r)}^2}}{1 + s\frac{\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2}} = 1 + \frac{rsf_t^2}{1 + sf_{t\times r}^2}$$

[37] W tym przypadku powinna być zastosowana formuła:

$$\hat{f}_t^2 = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_{s(t)}^2} = \frac{(MS_t - MS_{s(t)})/s}{MS_{s(t)}}$$

Na analogicznej zasadzie wyznaczona zostanie kappa dla efektu r:

$$\kappa_r = \frac{E(MS)_r}{E(MS)_{t\times r}} = \frac{rs\sigma_r^2 + s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}{s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}$$

A po podzieleniu licznika i mianownika przez wariancję wewnątrzgrupową i po dalszych przekształceniach otrzymamy:

$$\kappa_r = 1 + \frac{tsf_r^2}{1 + sf_{t\times r}^2}$$

Kappa dla efektu interakcji $t \times r$ zostanie wyznaczona według formuły:

$$\kappa_{t\times r} = \frac{E(MS)_{t\times r}}{E(MS)_{s(t\times r)}} = \frac{s\sigma_{t\times r}^2 + \sigma_{s(t\times r)}^2}{\sigma_{s(t\times r)}^2} = 1 + s\frac{\sigma_{t\times r}^2}{\sigma_{s(t\times r)}^2} = 1 + sf_{t\times r}^2$$

Model jednoczynnikowy z efektem stałym *T*

Przypomnijmy, że dla efektów stałych wyznaczamy lambdę. W modelu jednoczynnikowym parametr ten przybierze postać:

$$\lambda_T = n\frac{\Theta_T^2}{\sigma_{s(T)}^2} = n\frac{\theta_T^2\frac{(t-1)}{t}}{\sigma_{s(T)}^2} = n\frac{\frac{\sum\alpha_i^2}{(t-1)}\times\frac{(t-1)}{t}}{\sigma_{s(T)}^2} = n\frac{\frac{\sum\alpha_i^2}{t}}{\sigma_{s(T)}^2} = nf_T^2$$

Przypomnijmy też, że:

$$\frac{\sum\alpha_i^2}{t} = \frac{\sum(\mu_i - \mu)^2}{t}$$

Model dwuczynnikowy z efektami stałymi *T* oraz *R* w strukturze krzyżowej $T \times R$

W modelu dwuczynnikowym $T \times R$ parametry lambda dla trzech możliwych efektów stałych przybiorą postać odpowiednio:

$$\lambda_T = n\frac{\Theta_T^2}{\sigma_{s(T\times R)}^2} = n\frac{\frac{\sum\alpha_i^2}{t}}{\sigma_{s(T)}^2} = nf_T^2$$

$$\lambda_R = n\frac{\Theta_R^2}{\sigma_{s(T\times R)}^2} = n\frac{\frac{\sum \beta_j^2}{r}}{\sigma_{s(T)}^2} = nf_R^2$$

$$\lambda_{T\times R} = n\frac{\Theta_{T\times R}^2}{\sigma_{s(T\times R)}^2} = n\frac{\frac{\sum\sum(\alpha\beta)_{ij}^2}{rt}}{\sigma_{s(T\times R)}^2} = nf_{T\times R}^2$$

Przypomnijmy też, że:

$$\frac{\sum \alpha_i^2}{t} = \frac{\sum(\mu_i - \mu)^2}{t}$$

$$\frac{\sum \beta_j^2}{r} = \frac{\sum\left(\mu_j - \mu\right)^2}{r}$$

$$\frac{\sum\sum(\alpha\beta)_{ij}^2}{rt} = \frac{\sum\left(\mu_{ij} - \mu_i - \mu_j + \mu\right)^2}{rt}$$

Model dwuczynnikowy z efektem stałym *T* oraz efektem losowym *r* w strukturze krzyżowej $T \times r$

Model ten zawiera efekty mieszane – jeden efekt stały i dwa losowe. Dla efektu stałego wyznaczona będzie lambda:

$$\lambda_T = n\frac{\Theta_T^2}{s\sigma_{T\times r}^2 + \sigma_{s(T\times r)}^2} = \frac{n\frac{\frac{\sum \alpha_i^2}{t}}{\sigma_{s(T\times r)}^2}}{1 + s\frac{\sigma_{T\times r}^2}{\sigma_{s(T\times r)}^2}} = \frac{nf_T^2}{1 + sf_{T\times r}^2}$$

Dla efektów losowych wyznaczona będzie kappa:

$$\kappa_r = \frac{E(MS)_r}{E(MS)_{s(T\times r)}} = \frac{ts\sigma_r^2 + \sigma_{s(T\times r)}^2}{\sigma_{s(T\times r)}^2} = 1 + ts\frac{\sigma_r^2}{\sigma_{s(T\times r)}^2} = 1 + tsf_r^2$$

$$\kappa_{T\times r} = \frac{E(MS)_{T\times r}}{E(MS)_{s(T\times r)}} = \frac{s\sigma_{T\times r}^2 + \sigma_{s(T\times r)}^2}{\sigma_{s(T\times r)}^2} = 1 + s\frac{\sigma_{T\times r}^2}{\sigma_{s(T\times r)}^2} = 1 + sf_{T\times r}^2$$

Koele (1982: 515) zaznacza jednocześnie, że dokładne ustalenie poziomy mocy dla interakcji może być bardzo trudne z uwagi na obwaro-

wanie analiz surowymi założeniami dotyczącymi struktury macierzy kowariancji. Natomiast Jackson i Brashers (1994b: 372) przekonują, że problemy te nie dotyczą sytuacji, w której liczba poziomów czynnika stałego T wynosi $t = 2$.

W załączniku 1. aneksu Czytelnik znajdzie tabele (27–29) z wielkościami mocy dla efektu T oraz efektu interakcji $T \times r$. Wszystkie obliczenia wykonano przy założeniu, że $t = 2$.

Wracając jeszcze do treści paragrafu 1.2.3.3, zauważmy, że wzór na λ_T uwzględnia trzy z czterech wymienionych tam czynników wpływających na wielkość mocy testu efektu T. Są to: wielkość efektu czynnika T, wielkość efektu interakcji $T \times r$ oraz liczba osób badanych w grupie s (a także – jak pokazuje wzór – liczba osób w próbie n, która jest dodatkowo funkcją liczby grup t). Natomiast czwarty wymieniony tam czynnik – liczba replikacji – jest odzwierciedlony przez df_2.

Model dwuczynnikowy z efektem stałym T oraz efektem losowym r w strukturze zagnieżdżonej $r(T)$

Ten model także zawiera efekty mieszane, ale ze względu na strukturę, w jakiej pozostają ze sobą czynniki będzie miał jeden efekt stały i jeden efekt losowy. W tym modelu będą wyznaczone:

$$\lambda_T = n\frac{\Theta_T^2}{s\sigma_{r(T)}^2 + \sigma_{s(r(T))}^2} = \frac{n\dfrac{\sum \alpha_i^2}{t}}{\sigma_{s(r(T)}^2} = \frac{nf_T^2}{1 + sf_{r(T)}^2}$$

$$\kappa_{r(T)} = \frac{E(MS)_{r(T)}}{E(MS)_{s(r(T))}} = \frac{s\sigma_{r(T)}^2 + \sigma_{s(r(T))}^2}{\sigma_{s(r(T))}^2} = 1 + s\frac{\sigma_{r(T)}^2}{\sigma_{s(r(T))}^2} = 1 + sf_{r(T)}^2$$

W załączniku 2. aneksu Czytelnik znajdzie tabele (30–32) z wielkościami mocy dla efektu T oraz efektu $r(T)$. Wszystkie obliczenia wykonano przy założeniu, że $t = 2$.

W nawiązaniu do treści paragrafu 1.2.3.4. odnotujmy, że wzór na λ_T uwzględnia trzy z czterech wymienionych tam czynników, od których zależy wielkość mocy testu efektu T. Są to: wielkość efektu czynnika T, wielkość efektu czynnika r zagnieżdżonego w T oraz liczba osób badanych w grupie s (a także – zgodnie z wzorem – liczba osób w próbie n, która jest dodatkowo funkcją liczby grup t). Natomiast czwarty wymieniony tam czynnik – liczba replikacji – jest odzwierciedlony przez df_2.

2.7. Przykład analizy danych

Na koniec proponuję analizę danych dla modelu dwuczynnikowego z efektem stałym T i efektem losowym r w strukturze krzyżowej (dalej: model $T \times r$) oraz modelu dwuczynnikowego z efektem stałym T i efektem losowym r w strukturze zagnieżdżonej (dalej: model $r(T)$). Zestaw danych jest niewielki – został celowo zminiaturyzowany, tak aby mógł być zamieszczony w książce (załącznik 3. aneksu, tabela 33.), a Czytelnik – w razie potrzeby – mógł samodzielnie wykonać zaprezentowane niżej analizy. Dodatkowo dane zostały przygotowane w ten sposób, aby ten sam zestaw mógł być wykorzystany do modelu $T \times r$ oraz modelu $r(T)$. Zabieg ten zastosowano w celach poglądowych dla unaocznienia powiązań oraz różnic między tymi modelami. W nawiązaniu do rozdziału I zademonstrowano również, do jakich wyników i konkluzji (niekiedy niepoprawnych) doszedłby badacz, gdyby czynnik losowy r został przez niego w analizach 1) potraktowany jako stały; 2) pominięty.

Analizy przeprowadzono za pomocą IBM® SPSS® Statistics. W załączniku 4. aneksu Czytelnik znajdzie komendy (do wprowadzenia i uruchomienia w Edytorze Poleceń) za pomocą, których uzyskano wyniki zaprezentowane w tabelach 18–25.

Sytuacja problemowa, jaka będzie rozpatrywana, dotyczy eksperymentu sondażowego przeprowadzonego w celu udzielenia odpowiedzi na pytanie, czy styl prowadzenia wywiadu przez ankietera – socjoemocjonalny vs. formalny – wpływa na odpowiedzi respondentów[38], tutaj odnoszących się do pytań drażliwych. Styl prowadzenia wywiadu jest czynnikiem stałym, oznaczonym jako T, o poziomach t_1 (styl socjoemocjonalny) oraz t_2 (styl formalny). Badaną cechą jest skłonność respondentów do udzielania szczerych odpowiedzi. Odpowiada jej zmienna ilościowa *indeks*, która potencjalnie przyjmuje wartości od 0 do 30 (im wyższa wartość zmiennej, tym większa skłonność). Zbiór zawiera dane dla 30 respondentów (osoby badane są czynnikiem losowym oznaczonym jako s; liczba osób w jednej grupie wynosi 5).

Rozpocznijmy od **modelu $\boldsymbol{T \times r}$**. Do eksperymentu zaangażowano trzech ankieterów. Ankieterzy są czynnikiem losowym r, przyjmującym poziomy r_1, r_2 oraz r_3. Każdy ankieter zrealizował łącznie dziesięć wywiadów, pięć w stylu socjoemocjonalnym i pięć w stylu formalnym[39].

[38] Bezpośrednią inspirację do przykładu zaczerpnęłam od Dijkstry (1983), natomiast sama problematyka stylów i technik prowadzenia wywiadu przez ankietera ma w metodologii badań sondażowych długą tradycję i sięga m.in. prac Kahna i Cannella (1957), Converse i Schumana (1974), Sitka (1976).

[39] Liczba poziomów dla czynników: styl wywiadu, ankieter, respondent wynosi odpowiednio $t = 2$, $r = 3$, $s = 5$.

Tabela 18. przedstawia średnie na poszczególnych poziomach czynnika T, czynnika r oraz kombinacji poziomów tych czynników. Wyniki analizy wariancji przedstawia natomiast tabela 19.

Tabela 18. Średnie grupowe dla modelu $T \times r$

	r_1	r_2	r_3	Średnie w grupach czynnika T
t_1	14.6	21.6	18.4	18.2
t_2	17.8	14.4	12.2	14.8
Średnie w grupach czynnika r	16.2	18.0	15.3	–

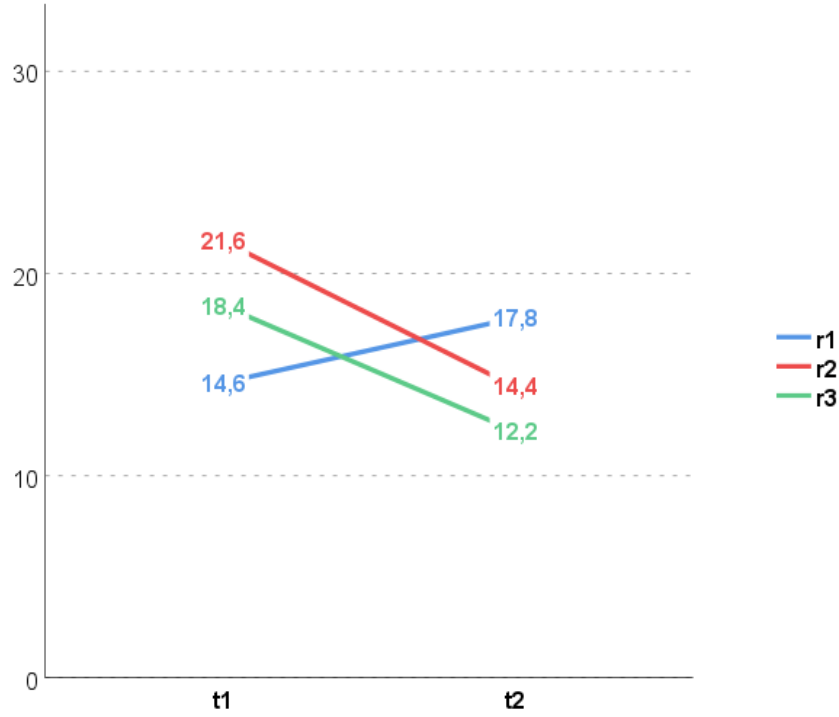
Źródło: opracowanie własne.

Tabela 19. Wyniki analizy wariancji dla modelu $T \times r$ (czynnik r losowy)

Źródło	SS	df	MS	F	wartość p
T	86.70	$t-1=1$	86.70	1.053	0.413
r	37.80	$r-1=2$	18.90	2.662	0.090
$T \times r$	164.60	$(t-1)\times(r-1)=2$	82.30	11.592	<0.001
$s(T \times r)$	170.40	$(s-1)\times t\times r=24$	7.10	–	–
Razem	459.50	29	–	–	–

Źródło: opracowanie własne.

W nawiązaniu do podrozdziału 2.2 oraz paragrafu 2.2.1 przypomnijmy, że w modelu $T \times r$, w którym ankieterów (r) traktujemy jako czynnik losowy, test efektu stylu wywiadu (T) musi dodatkowo uwzględniać losową zmienność pochodzącą z interakcji $T \times r$. Innymi słowy, jeżeli ustalenia dotyczące efektu T miałyby mieć szeroki zakres i odnosić się do nieprzebadanych respondentów i jednocześnie do nieprzebadanych ankieterów, to składnik błędu dla ilorazu F powinien uwzględnić zarówno zmienność pochodzącą od respondentów, jak i od interakcji $T \times r$ mówiącej o tym, czy efekt stylu jest taki sam dla każdego ankietera (zob. tabela 11.). Przy uwzględnieniu tej dodatkowej zmienności okazuje się, że efekt stylu T jest nieistotny statystycznie ($p = 0.413$), choć średnie indeksu sugerują, że styl socjoemocjonalny (18.2) bardziej sprzyja udzielaniu szczerych odpowiedzi niż styl formalny (14.8). Nieistotność efektu stylu wywiadu T jest pochodną faktu, że w analizowanym przykładzie zmienność pochodząca z interakcji jest bardzo duża. Żeby lepiej to uwidocznić, przedstawmy graficznie efekt interakcji, czyli efekt stylu dla poszczególnych ankieterów. Jak pokazuje wykres 1. efekt stylu wywiadu nie jest taki sam dla wszystkich ankieterów, a pamiętajmy przy tym, że trójkę przebadanych ankieterów traktujemy zaledwie jako próbkę.



Wykres 1. Efekt interakcji $T \times r$

Źródło: opracowanie własne.

Uzupełnijmy wyniki o oszacowania wielkości poszczególnych efektów (na podstawie wzorów dostarczonych w podrozdziale 2.5):

$$\widehat{\omega}_T^2 = 0.0063$$

$$\hat{\rho}_{I:r} = 0.0503$$

$$\hat{\rho}_{I:T\times r} = 0.6409$$

Wielkości te mówią, że ponad 64% całkowitej zmienności zmiennej *indeks* jest objaśniany przez interakcję ankietera i stylu wywiadu $T \times r$, a tylko 0.6% przez styl wywiadu T. Czynnik losowy ankieter r objaśnia 5% tej zmienności.

Porządkując otrzymane ustalenia – jedyny istotny efekt, jaki został uzyskany dotyczy interakcji $T \times r$. Gdyby czynniki tworzące interakcję były stałe, kolejnym krokiem byłaby analiza prostych efektów (analiza efektu jednego czynnika na poziomach drugiego czynnika). Tutaj jed-

nak – przez wzgląd na fakt, że ankieterzy są czynnikiem losowym – analiza efektu T na poziomach czynnika r nie jest zasadna z formalnego punktu widzenia, a także nie ma większego sensu merytorycznego. Żadna korzyść poznawcza dla nauki nie płynęłaby z ustalenia, że dla ankieterów Malinowskiego i Kowalskiego z Łodzi efekt stylu przybiera taką postać, że styl socjoemocjonalny sprzyja większej otwartości respondentów, a w przypadku ankietera Nowaka bardziej „sprawdza się" styl formalny (por. Jackson, Brashers, 1994a: 37).

Odnieśmy się jeszcze do efektu czynnika r. Różnice między ankieterami obserwowane po uśrednieniu ich wyników z wywiadów prowadzonych za pomocą obu stylów odzwierciedlają efekt ankietera r (średnie dla ankieterów widnieją w ostatnim wierszu tabeli 18.). Efekt ten nie jest istotny statystycznie, ale nawet wtedy gdyby był, kontynuacja analiz byłaby pozbawiona większego sensu, gdyż – podobnie jak wcześniej – byłoby to poszukiwanie różnic między konkretnymi ankieterami, co nie niosłoby ze sobą korzyści poznawczej, ale też nie pasowałby do zastosowanego tu podejścia, które nakazuje traktować zaangażowanych do eksperymentu ankieterów jako jedną z wielu możliwych próbek ankieterów (Jackson, Brashers, 1994a: 36).

Zobaczmy teraz, jakie rozstrzygnięcie w odniesieniu do efektu czynnika T uzyskalibyśmy, traktując czynnik ankieterzy jako stały R, a więc gdyby model przyjął postać $T \times R$. Zgodnie z wynikami w tabeli 20., efekt czynnika T zdecydowanie zyskał i jest teraz istotny statystycznie. Stało się tak za sprawą zmiany składnika błędu, który tym razem nie zawiera losowej zmienności pochodzącej od interakcji, a także za sprawą większej liczby stopni swobody, która to liczba obecnie jest m.in. pochodną liczby respondentów. A zatem w tym modelu odkrycie badawcze głosi, że styl socjoemocjonalny bardziej sprzyja udzielaniu szczerych odpowiedzi niż styl formalny – ustalenie to można uogólnić na szerszą populację respondentów, ale niestety prawidłowość ta musi być uznana za reakcję na trzech konkretnych ankieterów, którzy wzięli udział w eksperymencie. Korzyść poznawcza jest zatem zdecydowanie wątpliwa.

Tabela 20. Wyniki analizy wariancji dla modelu $T \times R$ (czynnik R stały)

Źródło	SS	df	MS	F	wartość p
T	86.70	1	86.70	12.211	0.002
R	37.80	2	18.90	2.662	0.090
$T \times R$	164.60	2	82.30	11.592	<0.001
$s(T \times R)$	170.40	24	7.10	–	–
Razem	459.50	29	–	–	–

Źródło: opracowanie własne.

W odniesieniu do interakcji wynik się nie zmienił – jest ona nadal istotna statystycznie. Dalsza analiza interakcji jest formalnie zasadna, bo zarówno T, jak i R są teraz czynnikami stałymi, ale zasadność merytoryczna jest nadal wątpliwa – czynnik ankieterzy „ze swej natury" stały jednak nie jest, a ustalenia badawcze względem trójki konkretnych ankieterów „nadal" nie przedstawiają sobą większej wartości.

Podajmy też wartości oszacowań wielkości poszczególnych efektów. Zmiana statusu czynnika ankieter z losowego na stały przynosi zmianę wartości tych oszacowań. W wyniku tej zmiany zmniejszyła się wielkość całkowitej zmienności do wyjaśnienia, „zyskał" efekt czynnika T, a „stracił" efekt interakcji $T \times R$:

$$\widehat{\omega}_T^2 = 0.1706$$

$$\widehat{\omega}_R^2 = 0.0506$$

$$\widehat{\omega}_{T\times R}^2 = 0.3223$$

Sprawdźmy także, jakie ustalenie w odniesieniu do efektu stylu wywiadu przyniósłby model, w którym czynnik ankieter byłby w ogóle pominięty, a więc gdyby analizy przeprowadzić dla modelu jednoczynnikowego z efektem stałym T[40]. Jak przedstawia to tabela 21., efekt T jest ponownie istotny statystycznie, choć w tym modelu wartość p jest nieco większa niż w modelu $T \times R$. Stało się tak z tego prostego powodu, że w obecnym modelu zmienność niewyjaśniona jest większa niż w modelu poprzednim ($SS_{s(T)}$ odpowiada tu sumie SS_R, $SS_{T\times R}$, $SS_{s(T\times R)}$). Wielkość efektu T jest niewiele mniejsza i wynosi $\widehat{\omega}_T^2 = 0.1552$.

Tabela 21. Wyniki analizy wariancji dla modelu T (czynnik r pominięty)

Źródło	SS	df	MS	F	wartość p
T	86.70	1	86.70	6.512	0.016
$s(T)$	372.80	28	13.31	–	–
Razem	459.50	29	–	–	–

Źródło: opracowanie własne.

Przywołując argumentację z paragrafu 1.2.3, przypomnijmy jednak, że w tej sytuacji problemowej posłużenie się modelem jednoczynnikowym grozi złamaniem założenia o niezależności pomiarów, ponieważ ignorowany jest fakt, że obserwacje są w rzeczywistości

[40] Liczba poziomów dla czynnika styl wywiadu wynosi $t = 2$, a dla czynnika „respondent" wynosi teraz $s = 15$.

pogrupowane (na 1 ankietera przypada wielu respondentów i ich wyniki mogą wykazywać zależność). Aby zapobiec temu ryzyku, czynnik losowy ankieter powinien być wprowadzony do analiz, a zatem powinien być przyjęty model $T \times r$ (gdy dane mają strukturę krzyżową) lub $r(T)$ (gdy dane mają strukturę zagnieżdżoną) (Judd, McClelland, Culhane, 1995; Kenny, Judd, 1986).

Aby posłużenie się modelem jednoczynnikowym T mogło być uznane za poprawne, badacz powinien ustalić wartość korelacji wewnątrzklasowej (wewnątrz grup) i wykazać, że błąd wynikający z pogrupowania obserwacji nie występuje (przykład analiz przedstawiają Kenny, Judd, 1986). Alternatywną metodę, choć przynoszącą co do zasady ten sam rezultat, zaproponował B. Winer (zob. podrozdział 2.8).

Przejdźmy teraz do **modelu $r(T)$**. Jak już zapowiedziano, do analizy zostaną wykorzystane te same dane, co poprzednio, natomiast uwzględniając strukturę zagnieżdżoną, w jakiej są teraz czynniki, będziemy rozpatrywać wariant, w którym jeden ankieter może prowadzić wywiady albo w stylu socjoemocjonalnym albo formalnym, a zatem rozważamy przypadek, w którym dane pochodzą od sześciu, a nie od trzech ankieterów[41], jak ostatnio.

Tabela 22. przedstawia średnie na poszczególnych poziomach czynnika T oraz czynnika $r(T)$. Natomiast wyniki analizy wariancji są przedstawione w tabeli 23.

Tabela 22. Średnie grupowe dla modelu $r(T)$

	r_1	r_2	r_3	r_4	r_5	r_6	Średnie w grupach czynnika T
t_1	14.6	21.6	18.4	–	–	–	18.2
t_2	–	–	–	17.8	14.4	12.2	14.8
Średnie w grupach czynnika $r(T)$	14.6	21.6	18.4	17.8	14.4	12.2	–

Źródło: opracowanie własne.

[41] Pomimo tego liczba poziomów dla poszczególnych czynników nie zmienia się w stosunku do schematu ze strukturą krzyżową. Ponieważ czynnik ankieter jest zagnieżdżony w czynniku „styl wywiadu", liczba poziomów dla czynnika „ankieter" wynosi $r = 3$. Liczba poziomów dla czynnika „styl wywiadu" wynosi $t = 2$, a dla czynnika „respondent" wynosi $s = 5$.

Tabela 23. Wyniki analizy wariancji dla modelu $r(T)$ (czynnik r losowy)

Źródło	SS	df	MS	F	wartość p
T	86.70	$t-1=1$	86.70	1.713	0.261
$r(T)$	202.40	$t\times(r-1)=4$	50.60	7.127	<0.001
$s(r(T))$	170.40	$(s-1)\times t\times r=24$	7.10	–	–
Razem	459.50	29	–	–	–

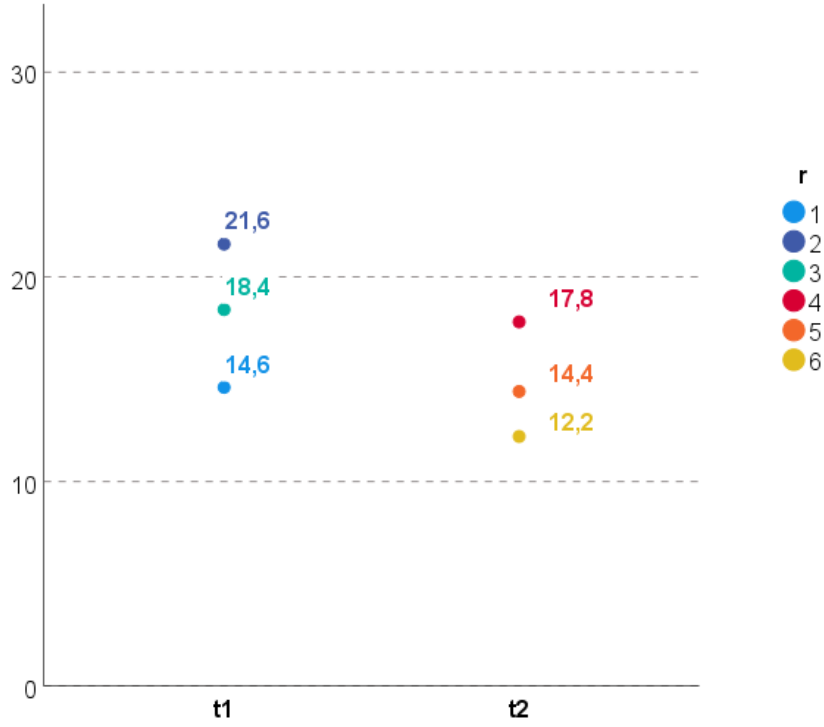
Źródło: opracowanie własne.

Jak widać w tabeli 23., gdy ankieterów uznamy jako źródło losowej zmienności, która – obok zmienności pochodzącej od respondentów – kształtuje wielkość średnich dla czynnika T w próbie (zob. tabela 12.), to efekt tego czynnika jest nieistotny statystycznie. Innymi słowy, gdyby ustalenia badawcze w odniesieniu do stylu prowadzenia wywiadu miały zostać rozszerzone na nieprzebadanych respondentów oraz nieprzebadanych ankieterów, to uzyskane w próbce dane nie pozwalają na konkluzję sugerowaną przez wartości średnich, że styl socjoemocjonalny zwiększa skłonność respondentów do udzielania szczerych odpowiedzi w porównaniu ze stylem formalnym. Nieistotność czynnika T jest w tym przypadku rezultatem zbyt dużych różnic w wynikach średnich, jakie osiągnęli ankieterzy prowadzący wywiady w stylu socjoemocjonalnym i/lub stylu formalnym (efekt $r(T)$ jest istotny statystycznie). Zróżnicowanie średnich wyników ankieterów prowadzących wywiady w stylu socjoemocjonalnym oraz ankieterów prowadzących wywiady w stylu formalnym uwyraźnia wykres, który jest graficzną ilustracją wartości średnich przedstawionych w tabeli 22.

W modelu tym uzyskano następujące szacunki wielkości efektów:

$$\widehat{\omega}_T^2 = 0.0708$$
$$\hat{\rho}_{I:r(T)} = 0.5117$$

Dla porządku podajmy jeszcze, jakie ustalenie w odniesieniu do efektu T przyniósłby model $R(T)$, a więc model, w którym ankieterzy uznani zostaliby za czynnik stały. Wynik przedstawia tabela 24. – jak widać rozstrzygnięcie dla efektu T jest takie samo, jakie przyniósł model $T\times R$ (składnikiem błędu jest w tym przypadku $MS_{s(R(T))}$, którego wartość jest taka sama jak $MS_{s(T\times R)}$). A zatem gdyby wnioski dotyczące stylu prowadzenia wywiadu miałyby być uogólnione na nieprzebadanych respondentów, ale jednocześnie byłyby ograniczone do szóstki ankieterów zaangażowanych w tym eksperymencie, to wyniki badania mówiłyby, że styl socjoemocjonalny sprzyja istotnie wyższym wartościom zmiennej *indeks* w porównaniu ze stylem formalnym. Taki zakres wniosków jest jednak – powtórzmy – niezadowalający.



Wykres 2. Efekt czynnika $r(T)$

Źródło: opracowanie własne.

Tabela 24. Wyniki analizy wariancji dla modelu $R(T)$ (czynnik R stały)

Źródło	SS	df	MS	F	wartość p
T	86.70	1	86.70	12.211	0.002
$R(T)$	202.40	4	50.60	7.127	<0.001
$s(R(T))$	170.40	24	7.10	–	–
Razem	459.50	29	–	–	–

Źródło: opracowanie własne.

Zwróćmy też uwagę na szacowane w tym modelu wielkości efektów oraz na ich powiązanie z wielkościami efektów w modelu $T \times R$. Tutaj te wielkości wynoszą:

$$\widehat{\omega}_T^2 = 0.1706$$

$$\widehat{\omega}_{R(T)}^2 = 0.3729$$

a zatem $\widehat{\omega}_T^2$ ma tę samą wartość w obu modelach, a $\widehat{\omega}_{R(T)}^2$ jest sumą wielkości $\widehat{\omega}_R^2$ oraz $\widehat{\omega}_{T\times R}^2$.

Analiza konkretnego przykładu danych jest dobrą okazją, by bliżej przyjrzeć się specyfice czynnika zagnieżdżonego. Tym razem będziemy abstrahować od tego, czy jest to czynnik losowy, czy stały, a więc tym czynnikiem może być zarówno $r(T)$, jak i $R(T)$.

W tym celu najlepiej będzie odnieść się do wzoru na $SS_{r(T)}$ (analogiczny wzór będzie dla $SS_{R(T)}$), by zobaczyć, jak operacyjnie zdefiniowana jest ta zmienność, o której mówimy, że jest zmiennością pochodzącą od ankieterów w obrębie stylów wywiadu (pochodzącą od replikacji r w obrębie poziomów czynnika T). Dodajmy też, że ten fragment tekstu jest bezpośrednim nawiązaniem do treści paragrafu 1.2.3.4.

Jak pokazuje poniższy wzór (Maxwell, Delaney, Kelly, 2018: 584; Keppel, Wickens, 2004: 558), $SS_{r(T)}$ jest zmiennością połączoną – sumą, na którą składa się zmienność wyników pracy ankieterów w grupie stosujących styl socjoemocjonalny oraz zmienność w grupie stosujących styl formalny.

$$SS_{r(T)} = \sum SS_{r\ na\ poziomie\ t_i} = SS_{r\ na\ poziomie\ t_1} + SS_{r\ na\ poziomie\ t_2}$$

$$df_{r(T)} = t \times (r-1)$$

Najprościej jest przekonać się o tym, w pierwszym kroku, dzieląc zbiór danych na dwa podzbiory według poziomów czynnika T (czyli wydzielając grupę stosujących styl t_1 oraz grupę stosujących styl t_2); w drugim kroku, wyznaczając wielkość SS_r w każdym podzbiorze; a w trzecim kroku, dodając te wielkości do siebie. Rezultat tych działań widoczny jest w tabeli 25. Daje ona wgląd w sumy kwadratów dla poszczególnych efektów przed i po dekompozycji efektu głównego $r(T)$ na efekty proste (efekt czynnika r na poziomie t_1 oraz efekt czynnika r na poziomie t_2).

Tabela 25. Sumy kwadratów odchyleń w modelu $r(T)$ wraz z dekompozycją zmienności dla czynnika $r(T)$

Źródło	SS	df	Źródło	SS	df
T	86.70	$t-1=1$	T	86.70	1
$r(T)$	202.40	$t\times(r-1)=4$	$r\ na\ poziomie\ t_1$	122.80	$(r-1)=2$
			$r\ na\ poziomie\ t_2$	79.60	$(r-1)=2$
$s(r(T))$	170.40	$(s-1)\times t\times r=24$	$s(r(T))$	170.40	24
Razem	459.50	29	Razem	459.50	29

Źródło: opracowanie własne.

Jak widać $SS_{r(T)} = 202.40 = 122.80 + 79.60$.

A zatem efekt $r(T)$ jest efektem ankietera, ale jednak specyficznie rozumianym, bo „oglądanym" oddzielnie w każdej grupie utworzonej przez czynnik T. Efekt $r(T)$ nie jest tożsamy z efektem r, który moglibyśmy nazwać „właściwym" efektem ankietera, a który udaje się ustalić tylko w schemacie ze strukturą krzyżową.

Pouczające w tym względzie będzie również zestawienie sum kwadratów, jakie uzyskano w modelu ze strukturą krzyżową i w modelu ze strukturą zagnieżdżoną dla dających się w tych modelach wyodrębnić źródeł. Wielkości te są przedstawione w tabeli 26.

Tabela 26. Sumy kwadratów i stopnie swobody dla poszczególnych źródeł w modelach ze strukturą krzyżową i zagnieżdżoną

Model ze strukturą krzyżową			Model ze strukturą zagnieżdżoną		
Źródła	SS	df	Źródła	SS	df
T	86.70	1	T	86.70	1
r	37.80	2	$r(T)$	202.40	4
$T \times r$	164.60	2			
$s(T \times r)$	170.40	24	$s(r(T))$	170.40	24
Razem	459.50	29	Razem	459.50	29

Źródło: opracowanie własne.

Tabela unaocznia, że czynnik $r(T)$ kumuluje w sobie dwie zmienności, które tylko w modelu ze strukturą krzyżową mogą być odseparowane: zmienność pochodzącą od czynnika r (którego efekt nazwano chwilę wcześniej „właściwym" efektem ankietera) oraz zmienność pochodzącą z interakcji $T \times r$. Regułę tę można ująć za pomocą formuły (Keppel, 1982: 224; Keppel, Wickens, 2004: 558):

$$SS_{r(T)} = \sum SS_{r\ na\ poziomie\ t_i} = SS_r + SS_{T\times r}$$

$$df_{r(T)} = df_r + df_{T\times r}$$

Jak widać $SS_{r(T)} = 202.40 = 37.80 + 164.60$.

Z punktu widzenia warunków do testowania efektu T w sytuacji, gdy r jest losowy, nie można określić z góry, który model tworzy bardziej korzystne warunki do odrzucenia hipotezy zerowej. Model ze strukturą krzyżową stwarza lepsze warunki, jeśli chodzi o porcję zmienności w mianowniku ilorazu F, która będzie mniejsza w porów-

naniu z porcją dostarczaną przez model ze strukturą zagnieżdżoną. Mocną stroną tego ostatniego modelu będzie jednak większa liczba stopni swobody, wpływająca nie tylko na wielkość MS w mianowniku, ale także na F_{kryt}.

2.8. Alternatywne strategie wobec modeli z czynnikami losowymi

Z włączeniem czynników losowych do modelu zwykle wiąże się niska moc testu dla efektów stałych. Dzieje się tak, ponieważ liczba replikacji jest zwykle mała, w każdym razie mniejsza niż liczba osób badanych, w efekcie do odrzucenia hipotezy zerowej o braku efektu czynnika stałego, potrzebne jest odpowiednio duże F – większe, gdy czynnik losowy jest obecny w modelu niż wtedy, gdy go nie ma. W celu demonstracji można posłużyć się danymi rozpatrywanymi w poprzednim podrozdziale. Testując efekt czynnika T na poziomie $alfa = 0.05$, to w modelu $T \times r$ wartość $F_{kryt}(df_1 = 1, df_2 = 2)$ wynosi 18.51, w modelu $T \times R$ wartość ta wynosi już tylko $F_{kryt}(df_1 = 1, df_2 = 24) = 4.25$, a modelu T (jednoczynnikowym) wynosi ona $F_{kryt}(df_1 = 1, df_2 = 28) = 4.19$. Mówiąc krótko, gdy w modelu obecne są czynniki losowe, szczególnie przy niewielkiej liczbie replikacji, trudniej jest wykazać istotny efekt głównej manipulacji eksperymentalnej. Mając ten problem na względzie, zaproponowano pewne procedury zaradcze.

Jedną ze strategii zaproponował B. Winer (1971: 378–384). Można ją nazwać **strategią upraszczania modelu**. Polega ona na usuwaniu z modelu tych źródeł stanowiących składniki błędu, które są nieistotne statystycznie na poziomie $alfa = 0.2$ (lub nawet większym $alfa = 0.3$) i mają mniej niż ok. 20 stopni swobody (df_2). Wyznaczając wysoki próg $alfa$, Winer chciał zapobiec pochopnemu usuwaniu z modelu tych źródeł, które same mają zbyt małą moc, by zostać wykryte[42].

Zauważmy, że strategii tej zastosować nie wolno w odniesieniu do żadnego z analizowanych w podrozdziale 2.7 modeli z efektami mieszanymi, czyli ani do modelu $T \times r$, ani $r(T)$. W obu przypadkach bowiem efekty stanowiące składniki błędu dla czynnika T są istotne statystycznie. W tej sytuacji redukcja któregoś modelu do postaci modelu jednoczynnikowego byłaby działaniem niepoprawnym.

[42] Jest to więc prosta alternatywa dla bardziej wymagającego zalecenia Brashers i Jackson (1999: 464–465), by poprzez odpowiednio dużą liczbę replikacji oraz osób badanych zapewnić wysoką moc efektom będącym składnikami błędu (zob. przypis 11. w paragrafie 1.2.3.3).

Dodajmy też za Winerem, że strategia upraszczania modelu – nawet przy spełnieniu powyższego warunku – ma swoich oponentów, którzy z zasady przeciwni są usuwaniu z modelu źródeł nieistotnych statystycznie. Przykładowo Jackson i Brashers nie promują w swoim piśmiennictwie tej strategii.

Inną metodę zaproponowali Glass i Hopkins (1996: 556–560). Jest to **strategia zwiększania poziomów generalizacji** (ang. *the incremental generalization strategy*). Opiera się ona na pomyśle, że w schematach przewidujących co najmniej jeden czynnik losowy uogólnianie można poddać gradacji. Przedstawmy tę strategię na przykładzie badania analizowanego w podrozdziale 2.7. W badaniu tym są dwa czynniki losowe – „badane osoby" oraz „ankieterzy". W tym przypadku można wyróżnić trzy poziomy generalizacji wniosku dotyczącego efektu stałego T:

- poziom zero – wniosek jest ograniczony do przebadanych osób oraz przebadanych ankieterów. W tej sytuacji model zawiera tylko czynniki stałe (czynnikiem stałym są również osoby badane), a to oznacza, że procedura wnioskowania statystycznego w ogóle nie jest uruchamiana i w ocenie efektu czynnika T brane są pod uwagę tylko statystyki opisowe.
- poziom pierwszy – w sytuacji odrzucenia hipotezy zerowej wniosek jest uogólniany na nieprzebadane osoby, ale ograniczony do przebadanych ankieterów. W tej sytuacji jedynym czynnikiem losowym są osoby badane (ankieterzy są czynnikiem stałym). W zależności od tego, czy dane zbierane są za pomocą schematu w strukturze krzyżowej czy zagnieżdżonej, model analityczny przyjmie postać – odpowiednio – $T \times R$ lub $R(T)$.
- Poziom drugi – w sytuacji odrzucenia hipotezy zerowej wniosek jest uogólniany na nieprzebadane osoby oraz nieprzebadanych ankieterów. W tym przypadku czynnikami losowymi są uczestnicy eksperymentu oraz ankieterzy. Model analityczny przyjmuje postać $T \times r$ lub $r(T)$.

Procedura wymaga rozpoczęcia analiz od przetestowania efektu T na poziomie pierwszym. Jeżeli na tym niższym poziomie efekt okazałby się nieistotny statystycznie, analiza zostaje zakończona. W sytuacji, w której efekt przeszedłby pomyślnie próbę testu, czyli okazałby się istotny statystycznie, analiza przenoszona jest na poziom wyższy. Jeżeli i na tym poziomie efekt przechodzi test pomyślnie, to wniosek osiąga najwyższy (drugi) poziom generalizacji, a jeśli testu tego nie przechodzi, to wniosek zostaje na pierwszym poziomie generalizacji.

Jak więc widać, metoda Glassa i Hopkinsa stwarza szansę na dostrzeżenie efektu T, o ile ten występuje na niższym poziomie generali-

zacji. Gdyby jedynym podejściem było podejście restrykcyjne (test tylko na poziomie drugim), występujący na poziomie pierwszym efekt T nie zostałby dostrzeżony. Metoda ta bierze zatem pod uwagę to, że niekiedy brak efektu na wyższym poziomie generalizacji może być wynikiem zbyt małej mocy testu, a nie prawdziwości hipotezy zerowej.

Gdyby wykorzystać tę strategię do analizowanego przykładu danych, to okazałaby się owocna – przypomnijmy, że w modelach $T \times R$ oraz $R(T)$, a więc na pierwszym poziomie generalizacji, efekt czynnika T był istotny statystycznie. Stosując zatem tę strategię, dostalibyśmy jako badacze pewien argument na rzecz prawdziwości hipotezy alternatywnej.

Jak przekonują autorzy jest też inna korzyść ze stosowania ich strategii, pod warunkiem przyjęcia poziomów generalizacji w zaproponowanym kształcie. Jest ona związana z możliwością wystąpienia nietypowego układu danych, w którym test efektu T przynosi wynik istotny statystycznie, gdy replikacje są czynnikiem losowym (drugi poziom generalizacji), natomiast gdy replikacje są czynnikiem stałym (pierwszy poziom generalizacji), to wynik jest nieistotny. Dochodzi zatem do paradoksu – wniosek dotyczący efektu T można rozszerzyć na wszystkie replikacje (tu: wszystkich ankieterów), ale jednocześnie nie można go zastosować do części replikacji (tu: przebadanych ankieterów). Zastosowanie się do strategii autorów miałoby zapobiec wystąpieniu tej nielogiczności, bowiem analizy zostałyby zatrzymane na pierwszym poziomie generalizacji. Propozycja broni się o tyle, że – jak przekonuje Hopkins (1983) – ów układ danych świadczy o braku niezależności pomiarów, a więc zasadniczo o ich wadliwości. W modelu $r(T)$ chodziłoby o taką sytuację, w której zmienność pochodząca od czynnika $r(T)$ (np. ankieterów) jest „nienaturalnie" mała i dochodzi do tego, że testując efekt T, podstawy do odrzucenia hipotezy zerowej są mocniejsze, gdy składnikiem błędu jest $MS_{r(T)}$ niż gdy jest nim $MS_{s(R(T))}$. W komentarzu można zwrócić uwagę, że w przywoływanym przez Hopkinsa (1983) nietypowym badaniu, w którym zastosowano model $r(T)$, liczba replikacji była bardzo mała (3 replikacje na 1 poziom czynnika T), a poziom istotności bardzo wyśrubowany $alfa = 0.01$ (przy poziomie 0.05 efekt T jest istotny statystycznie na pierwszym i drugim poziomie generalizacji). Być może więc problem z danymi umiejscowiony jest gdzie indziej – schemat nie zapewnił odpowiedniej mocy testu czynnika $r(T)$, a restrykcyjny poziom istotności tylko ten problem pogłębił.

Inna uwaga, jaką można wysunąć wobec tej strategii dotyczy konstrukcji poziomów generalizacji. Propozycja Glassa i Hopkinsa przewiduje, że proces uogólniania należy rozpocząć od uczestników. Prawdo-

podobnie w tym punkcie propozycja ta odzwierciedla rozumowanie większości badaczy na temat kierunku zwiększania generalizacji. Trzeba jednak zaznaczyć, że nie jest to jedyne możliwe rozwiązanie. Przywoływany już wcześniej Clark (1973) porównywał wyniki trzech modeli, z których tylko dwa pierwsze pasują do „skali" i odpowiadają odpowiednio pierwszemu i drugiemu poziomowi generalizacji (1/ osoby – losowy, replikacje – stały, 2/ osoby – losowy, replikacje – losowy), natomiast trzeci, uprawniony przecież model (3/ osoby – stały, replikacje – losowy) w tej propozycji się nie mieści.

Czytelnikom zainteresowanym tym, w jaki sposób łączyć strategię zwiększania poziomów generalizacji ze strategią upraszczania modelu, można polecić lekturę artykułu Hopkinsa (1983).

*

W podsumowaniu rozdziału II należy podkreślić, że kluczową umiejętnością w analizie wariancji dla modeli zawierających czynniki losowe jest umiejętność wyznaczania oczekiwanych średnich kwadratów $E(MS)$. Jest to „okno na świat", które pozwala dokonać „rozbioru" modelu i dobrać właściwy składnik błędu do przetestowania każdego efektu nie tylko w modelach międzygrupowych (dla grup niezależnych), ale także wewnątrzgrupowych (dla grup zależnych) i mieszanych. Dzięki tej umiejętności badacz zyskuje bardzo potrzebną kontrolę nad działaniami oprogramowania statystycznego. W połączeniu ze znajomością składni poleceń danego oprogramowania badacz będzie też potrafił dokonać ewentualnych korekt tych działań. Umiejętność wyznaczania $E(MS)$ jest także punktem wyjścia innych procedur analitycznych, takich jak ustalanie postaci formuł pozwalających wyznaczyć szacunki komponentów wariancyjnych, wielkość poszczególnych efektów w konkretnych modelach, czy też pozwalających wyznaczyć moc testów (przy pewnych zastrzeżeniach).

Warto w tym miejscu wskazać ograniczenia tego opracowania i polecić Czytelnikom literaturę do dalszego studiowania.

Rozpatrywane tu przykłady danych przewidywały najprostszą sytuację, w której – obok czynnika losowego – występował tylko jeden czynnik stały (T), a do tego przyjmował on tylko dwa poziomy: t_1 i t_2. Tym samym w odniesieniu do czynnika stałego wywód został ograniczony do testowania efektów głównych – w opracowaniu tym nie omówiono testów efektów prostych (gdyby czynnik stały miał więcej niż dwa poziomy), ani testów prostych efektów interakcyjnych (gdyby w modelu wystąpił drugi czynnik stały, będący z tym pierwszym w strukturze krzyżowej). Czytelnikom zainteresowanym testowaniem

tych efektów polecam pozycję Keppela i Wickensa (2004) a także Sahai, Ageel (2000). Przy okazji można dodać, że rozwiązania w tym zakresie będą opierały się na znajomości $E(MS)$.

W prezentowanym opracowaniu nie poświęcono praktycznie miejsca modelom czysto losowym, tj. zawierającym jedynie czynniki losowe. Czytelnikom zainteresowanym tymi modelami, ich dokładniejszą charakterystyką, przykładami oraz procedurami analitycznymi, w tym przede wszystkim analizą komponentów wariancyjnych, można polecić pozycję Stanisza (2007), a następnie Sahai, Ageel (2000) oraz Sahai, Ojeda (2004). Z analizami tego rodzaju, ale prowadzonymi w perspektywie teorii uniwersalizacji Czytelnik może zapoznać się w pracach: Shavelson, Webb (1991), Brennan (2001).

Sposoby wyznaczania przedziałów ufności dla parametrów, a także procedury analityczne dla schematów niezbalansowanych (o nierównolicznych grupach) przedstawiają Sahai, Ageel (2000).

Czytelnikom zainteresowanym schematami wewnątrzgrupowymi warto polecić podręcznik Keppela i Wickensa (2004).

Dla użytkowników pakietu SAS i SPSS pomocny będzie podręcznik Sahai, Ageel (2000), a dla użytkowników pakietu STATISTICA PL – książka Stanisza (2007).

Zakończenie

W książce starałam się uzasadnić, że wiele pytań problemowych, które w naukach społecznych rozstrzygamy na drodze eksperymentu, wymaga wzbogacenia planu badawczego o replikacje i uwzględnienia ich w analizach jako czynnika losowego. Postulat ten wynika ze specyfiki cech i zjawisk, których oddziaływanie chcemy badać, a których często nie daje się przełożyć na bodziec zawierający daną cechę w czystej, wyizolowanej postaci. Podjęte tutaj zagadnienie należy do obszaru metodologii badań eksperymentalnych, jako że podnosi kwestię planowania eksperymentów w sposób mający zapewnić ustaleniom badawczym wewnętrzną i zewnętrzną trafność. Ale zagadnienie to ma też aspekt statystyczny – analiza zgromadzonych danych wymaga zastosowania metod ilościowych, a obecność czynników losowych w modelu, czyni te analizy dodatkowo złożonymi. Celem moim było, by to zagadnienie naświetlić z obu stron, wypełniając w ten sposób lukę w polskim piśmiennictwie, w którym nie ma pozycji łączącej aspekt planowania eksperymentów uwzględniających czynniki losowe z aspektem analizowania otrzymanych danych.

Zaprezentowany w książce punkt widzenia na poprawnie zaplanowany eksperyment bierze swój początek od dyskusji toczonej przez badaczy komunikacji społecznej. Odniesiono się tam krytycznie do dużej części przeprowadzonych wcześniej badań eksperymentalnych. Wskazano, że ich wadliwe zaprojektowanie sprzyja „odkrywaniu" prawidłowości rzekomych – takich, które faktycznie nie istnieją. Możliwość zaradzenia tym ułomnościom dostrzeżono w czynnikach losowych i ich potencjale. Jednak akceptacja tego remedium nie była w pełni oczywista, wymagała bowiem zliberalizowania rygorystycznego stanowiska, że za losowy może być uznany jedynie czynnik, którego poziomy są losowo wybrane.

Na koniec warto więc przyjrzeć się zmianom praktyk badawczych, jakie zaszły w dziedzinie badań nad komunikacją społeczną. W jakim zakresie nowe rozwiązania przyjęły się na gruncie, z którego wyszedł „rewolucyjny" impuls? Dokonując oceny eksperymentów, których wyniki zostały opublikowane w czasopiśmie *Human Commmunication Research* w latach 1974–1998, Brashers i Jackson (1999) odnotowali, że 35% eksperymentów powinno wykorzystać replikacje jako czynnik losowy i wymóg ten spełniło 58% z tej grupy. Wyniki tych badań zapo-

wiadały też tendencję wzrostową. Biorąc pod uwagę czas ukazania się artykułów – „kamieni węgielnych" – Coleman (1964), Clark (1973), Jackson i Jacobs (1983) – nasuwa się konstatacja, że nowe rozwiązania przyjmowane są powoli, szczególnie gdy pozostają przedmiotem mniejszych bądź większych kontrowersji metodologicznych, a ich przyjęcie odbywa się kosztem porzucenia puryzmu w jednym punkcie na rzecz zwiększenia rygoryzmu w innym.

Moim zdaniem, „rewolucja" ta warta jest swojej ceny, choć być może wiąże się z ryzykiem dobierania prób w sposób niewystarczająco staranny i bez należytego zrozumienia, że ideałem pozostaje próba losowa, a rezygnacja z tego sposobu doboru wymaga uzasadnienia. Natomiast skupiając się na korzyściach, chciałabym wyeksponować jeszcze jedną z nich, uprzednio niewskazywaną. Obecność replikacji jako czynnika losowego w modelu nie tylko zwiększa trafność ustaleń w odniesieniu do efektu *T* oraz pozwala ocenić zmienność, jaka towarzyszy temu efektowi, ale umożliwia także nieco inne podejście do samego efektu, właśnie jako mającego zmienną i zależną od okoliczności „naturę". Innymi słowy, tam, gdzie zależność efektu od replikacji jest duża, a replikacje zróżnicowane i różnice między nimi uchwytne, powinniśmy nie tylko być zainteresowani „uśrednioną" wartością efektu i zakończyć analizy konstatacją, że jest on ogólnie mały (bo pewnie taki w tej sytuacji będzie), ale także zwrócić uwagę na to, jakie okoliczności sprzyjają jakim wartościom średnim (Brashers, Jackson, 1999). Potencjał eksperymentów z replikacjami polega na wzmocnieniu konkluzywności wniosków z badań. Wydają się one – na pewno częściowym, ale istotnym – remedium na sytuację, w której pojedyncze badania prowadzone na ten sam temat przynoszą niespójne wyniki i domagają się odroczonych w czasie metaanaliz.

Na sam koniec dodajmy, że i w obrębie samej metodologii metaanaliz rozważa się, w jaki sposób traktować zbiór badań poddawanych metaanalizie – jako czynnik stały, czy losowy (Rószkiewicz i in., 2013). A zatem dyskusja w tej newralgicznej kwestii toczy się nie tylko na poziomie badań pierwotnych, ale także wtórnych.

Bibliografia

Aczel A.D. (2000), *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa.

Ato M., Vallejo G., Palmer A. (2013), *The two-way mixed models: A long and winding controversy*, „Psicothema", vol. 25, no. 1, s. 130–136.

Boik R.J. (2005), *Randomization*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science,* vol. 4, John Wiley & Sons, Chichester.

Bonge D.R., Schuldt W.J., Harper Y.Y. (1992), *The experimenter-as-fixed-effect fallacy*, „The Journal of Psychology", vol. 126, no. 5, s. 477–486.

Brashers D.E., Jackson S. (1999), *Changing conceptions of „message effects". A 24-year overview*, „Human Communication Research", vol. 25, no. 4, s. 457–477.

Brennan R.L. (2001), *Generalizability theory*, Springer Verlag, Berlin-Heidelberg.

Brown H., Prescott R. (2015), *Applied mixed models in medicine*. Third edition, John Wiley & Sons, Chichester.

Brzeziński J.M. (2008), *Badania eksperymentalne w psychologii i pedagogice*, Wydawnictwo Naukowe Scholar, Warszawa.

Brzeziński J.M. (2009), *Kiedy odwołując się do testów psychologicznych postępujemy nieetycznie? Analiza kontekstu*, „Czasopismo Psychologiczne", t. 15, nr 2, s. 321–332.

Clark H.H. (1973), *The language-as-fixed-effect fallacy: A critique of language statistics in psychological research*, „Journal of Verbal Learning and Verbal Behavior", vol. 12, s. 335–359.

Clark H.H. (1976), *Reply to Wike and Church*, „Journal of Verbal Learning and Verbal Behavior", vol. 15, s. 257–261.

Cohen J. (1976), *Random means random*, „Journal of Verbal Learning and Verbal Behavior", vol. 15, s. 261–262.

Cohen J. (1988), *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum Associates, New York, NY.

Coleman E.B. (1964), *Generalizing to language population*, „Psychological Reports", vol. 14, s. 219–226.

Coleman E.B. (1972–73), *Generalization variable and restricted hypotheses*, „Journal of Reading Behaviour", vol. 5, no. 4, s. 226–236.

Coleman E.B. (1979), *Generalization effects vs. random effects: Is σ^2_{TL} a source of type 1 or type 2 error?,* „Journal of Verbal Learning and Verbal Behavior", vol. 18, s. 243–256.

Converse J.M., Schuman H. (1974), *Conversations at random: survey research as interviewers see it*, John Wiley & Sons, New York, NY.

Cook T.D., Campbell D.T. (1979), *Quasi-experimentation. Design & analysis issues for field settings*, Houghton Mifflin Company, Boston, MA.

Cronbach L.J., Gleser G.C., Nanda H., Rajaratnam N. (1972), *The dependability of behavioral measurements. Theory of generalizability for scores and profiles*, Wiley, New York, NY.

Dijkstra W. (1983), *How interviewer variance can bias the results of research on interview effect?*, „Quality and Quantity", vol. 17, s. 179–187.

Field A.P. (2005), *Intraclass correlation*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science,* vol. 2, John Wiley & Sons, Chichester.

Forster K.I., Dickinson R.G. (1976), *More on the language-as-fixed-effect fallacy: Monte Carlo estimates or error rates for F_1, F_2, F', and min F'*, „Journal of Verbal Learning and Verbal Behavior", vol. 15, s. 135–142.

Fowler F.J., Mangione T.W. (1990), *Standardized survey interviewing. Minimizing interviewer-related error*, Applied Social Research Methods Series, vol. 18, Sage Publications, Newbury Park.

Fuchs M. (2009), Gender-of-interviewer effects in a video-enhanced Web survey, „Social Psychology", vol. 40, no. 1, s. 37–42.

Gamst G., Meyers L.S., Guarino A.J. (2008), *Analysis of variance designs. A conceptual and computational approach with SPSS and SAS,* Cambridge University Press, New York, NY.

Glass V.G., Hopkins K.D. (1996), *Statistical methods in education and psychology. Third edition*, Allyn and Bacon, Boston, MA.

Groves R.M. (1989), *Survey errors and survey costs*, John Wiley & Sons, New York, NY.

Hader R.J. (1973), *An improper method of randomization in exeperimental design*, „The American Statistician", vol. 27, s. 82–84.

Hedges L.V., Hedberg E.C. (2007), *Interclass correlation values for planning group randomized trials in education*, „Educational Evaluation and Policy Analysis", vol. 29, no. 1, s. 60–87.

Hopkins K.D. (1976), *A simplified method for determining for expected mean sqaures and error terms in the analysis of variance*, „The Journal of Experimental Education", vol. 45, no. 2, s. 13–18.

Hopkins K.D. (1983), *A strategy for analyzing ANOVA designs having one or more random factors*, „Educational and Psycholigcal Measurement", vol. 43, s. 107–113.

Howell D.C. (2007), *Statistical methods for psychology. Seventh edition*, Wadsworth Cengage Learning, Belmont, CA.

Jabkowski P. (2015), *Reprezentatywność badań reprezentatywnych. Analiza wybranych problemów metodologicznych oraz praktycznych w paradygmacie całkowitego błędu pomiaru*, Seria Socjologia, nr 77, Wydawnictwo Naukowe UAM, Poznań.

Jackson S. (1991), *Meta-analysis for primary and secondary data analysis: the super-experiment metaphor*, „Commmunication Monographs", vol. 58, s. 449–462.

Jackson S. (1992), *Message effects research. Principles of design and analysis*, The Guilford Press, New York, NY.
Jackson S., Brashers D.E. (1994a), *Random factors in ANOVA*, Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07–098, Sage, Thousand Oaks, CA.
Jackson S., Brashers D.E. (1994b), *M>1, Analysis of treatment x replication design*, „Human Communication Research", vol. 20, no. 3, s. 356–389.
Jackson S., Jacobs S. (1983), *Generalizing about messages: suggestions for design and analysis of experiment*, „Human Communication Research", vol. 9, no. 2, s. 169–191.
Jackson S., Jacobs S. (1987), *The search for systematic message effects. Contributions of meta-analysis and better design*. Paper presented at the meeting of International Communication Association, Montreal.
Jackson S., Brashers D.E., Massey J. E. (1992), *Statistical testing in treatment by replication design: three options reconsidered*, „Communication Quarterly", s. 211–227.
Jackson S., O'Keefe D.J., Brashers D.E. (1994), *The messages replication factor: Methods tailored to messages as objects of study*, „Journalism Quarterly", vol. 71, no. 4, s. 984–996.
Jackson S., O'Keefe D.J., Jacobs S. (1988), *The search for reliable generalizations about messages. A comparison of research strategies*, „Human Communication Research", vol. 15, no. 1, s. 127–142.
Jackson S., O'Keefe D.J., Jacobs S., Brashers D. (1989), *Messages as replications: Toward a message-centered design starategy*, „Communication Monographs", vol. 56, no. 4, s. 364–384.
Judd C.M., McClelland G.H., Culhane S.E. (1995), *Data analysis: continuing issues in the everyday analysis of psychological data*, „Annual Review of Psychology", vol. 46, s. 433–465.
Kahn R.L., Cannel C.F. (1957), *The dynamics of interviewing*, John Wiley & Sons, New York, NY.
Kay E.J., Richter M.L. (1977), *The category-confound: A design error*, „The Journal of Social Psychology", vol. 103, s. 57–63.
Kenny D.A., Judd C.M. (1986), *Consequances of violeting the independence assumption in analysis of variance*, „Psychological Bulletin", vol. 99, no. 3, s. 422–431.
Keppel G. (1976), *Words as random variables*, „Journal of Verbal Learning and Verbal Behavior", vol. 15, s. 263–265.
Keppel G. (1982), *Design and analysis: a researcher's handbook. Second edition,* Prentice Hall, Englewood Cliffs, NJ.
Keppel G., Wickens T.D. (2004), *Design and analysis: a researcher's handbook. Fourth edition*, Pearson Prentice Hall, Upper Saddle River, NJ.
Kirk R.E. (1968), *Experimental design: procedures for the behavioral sciences*, Brooks/Cole Publishing Company, Belmont, CA.
Kirk R.E. (1996), *Practical significance: a concept whose time has come*, „Educational and Psycholigical Measurement", vol. 56, no. 5, s. 746–759.

Kirk R.E. (2005), *Effect size measures*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science,* vol. 2, John Wiley & Sons, Chichester.

Koele P. (1982), *Calculating power in analysis of variance*, „Psychological Bullettin", vol. 92, no. 2, s. 513–526.

Landis J.R., Sullivan J., Sheley (1973), *Feminist attitudes as related to sex of the interviewer*, „Pacific Sociological Review", vol. 16, s. 305–314.

Lindman H.R. (1992), *Analysis of variance in experimental design*, Springer-Verlag, New York, NY.

Lisowska-Magdziarz M. (2004), *Analiza zawartości mediów. Przewodnik dla studentów*, seria wydawnicza: Zeszyty Wydziałowe, Zeszyt nr 1, Uniwersytet Jagielloński.

Malarska A. (2005), *Statystyczna analiza danych wspomagana programem SPSS*, SPSS Polska, Kraków.

Martindale C. (1978), *The therapist-as-fixed-effect fallacy in psychotherapy research*, „Journal of Consulting and Clinic Psychology", vol. 46, no. 6, s. 1526–1530.

Maxwell S. E., Delaney H.D., Kelly K. (2018), *Designing experiments and analyzing data. A model comparison perspective. Third edition*, Routledge, Taylor & Francis, New York, NY.

Mcgraw K.O., Wong S.P. (1996), *Forming inferences about intraclass correlation coefficients*, „Psychological Methods", vol. 1, no. 1, s. 30–46.

Pilkonis P.A., Imber S.D., Lewis P., Rubinsky P. (1984), *A comparative outcome study of individual, group, and conjoint psychotherapy*, „Psychotherapy", vol. 41, s. 431–437.

Richter M.L., Seay M.B. (1987), *ANOVA designs with subjects and stimuli as random effects: Applications to prototype effects on recognition memory*, „Journal of Personality and Social Psychology", vol. 53, no. 3, s. 470–480.

Rosenthal R., Rosnow R.L. (2009), *Artifacts in behavioral research. Robert Rosenthal and Ralph L. Rosnow's Classis Books*, Oxford University Press, New York, NY.

Rozmus D. (2011), *Analiza wariancji*, [w]: E. Gatnar, M. Walesiak (red.), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, Wydawnictwo C.H. Beck, Warszawa.

Rószkiewicz M., Perek-Białas J., Węziak-Białowolska D., Zięba-Pietrzak A. (2013), *Projektowanie badań społeczno-ekonomicznych. Rekomendacje i praktyka badawcza*, Wydawnictwo Naukowe PWN, Warszawa.

Rumenik D.K., Capasso D.R., Hendrick C. (1977), *Experimenter sex effects in behavioral research*, „Psychological Bulletin", vol. 84, no. 5, s. 852–877.

Sahai H., Ageel M.I. (2000), *The analysis of variance. Fixed, random and mixed models*, Birkhäuser, Boston, MA.

Sahai H., Ojeda M.M. (2004), *The analysis of variance for random models: theory, methods, applications and data analysis. Volume 1: Balance data*, Springer Science+Business Media, New York, NY.

Santa J.L., Miller J.L., Shaw M.L. (1979), *Using quasi F to prevent alpha inflation due to stimulus variation*, „ Psychological Bulletin", vol. 86, no. 1, s. 37–46.

Satterthwaite F.E. (1946), *An approximate distribution of estimates of variance components*, „Biometrics Bulletin", vol. 2, s. 110–114.

Scheffé H.A. (1959), *The analysis of variance*, John Wiley & Sons, New York, NY.

Schwarz C.J. (1993), *The mixed-model ANOVA: The truth, the computer packages, the book. Part I: Balanced data*, „The American Statistican", vol. 47, no. 1, s. 48–59.

Shadish W.R., Cook T.D., Campbell D.T. (2002), *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin Company, Boston, MA.

Shavelson R.J., Webb N.M. (1991), *Generalizability theory. A primer,* Sage, Newbury Park.

Sitek W. (1976), *Ankieter humanista i ankieter technik. Teoretyczne i praktyczne aspekty typologicznej selekcji ankieterów*, [w]: Z. Gostkowski (red.), *Z metodologii i metodyki socjologicznych badań terenowych:* zeszyt 4, Warszawa, IFiS PAN, s. 63–113.

Smith P.T. (2005), *Random effects and fixed effects fallacy*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science,* vol. 4, John Wiley & Sons, Chichester.

Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 2. Modele liniowe i nieliniowe*, StatSoft Polska, Kraków.

Sułek A. (1979), *Eksperyment w badaniach społecznych*, PWN, Warszawa.

Sułek A. (2002), *Ogród metodologii socjologicznej*, Wydawnictwo Naukowe Scholar, Warszawa.

Sztabiński P.B. (1997), *Ankieterzy i ich respondenci. Od kogo zależą wyniki badań ankietowych*, IFiS PAN, Warszawa.

Szymczak W. (2018), *Podstawy statystyki dla psychologów. Podręcznik akademicki*, Difin, Warszawa.

Tchach C., Berger V.W. (2005), *Simple random assignment*, [w:] B.S. Everitt, D.C. Howell (red.), *Encyclopedia of Statistics in Behavioral Science*, vol. 4, John Wiley & Sons, Chichester.

Tversky A., Kahneman D. (1971), *Belief in the law of small numbers*, „Psychological Bulletin", vol. 76, no. 2.

Vaughan G.M., Corballis M.C. (1969), *Beyond tests of significance: Estimating strength of effects in selected ANOVA designs*, „Psychological Bulletin", vol. 72, no. 3, s. 204–213.

Vogt W.P., Johnson R.B. (2016), *The Sage dictionary of statistics & methodology. A nontechnical guide for the social sciences. Fifth edition*, Sage, Los Angeles, CA.

Wells G.L., Windschitl P.D. (1999), *Stimulus sampling and social psychological experimentation*, „Personality and Social Psychology Bulletin", vol. 25, s. 1115–1125.

West S.G. (2006), *Poza eksperyment laboratoryjny – plany eksperymentalne oraz quasi-eksperymentalne w otoczeniu naturalnym*, [w]: J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów*, Wydawnictwo Naukowe PWN, Warszawa.

Wickens T.D., Keppel G. (1983), *One the choice of design and of test statistic in the analysis of experiments with sampled materials*, „Journal of Verbal Learning and Verbal Behavior", vol. 22, s. 296–309.

Wike E.L., Church J.D. (1976), *Comments on Clark's "The Language-as-fixed-effect fallacy"*, „Journal of Verbal Learning and Verbal Behavior", vol. 15, s. 249–255.

Wiktorowicz J., Grzelak M.M., Grzeszkiewicz-Radulska K. (2020), *Analiza statystyczna z IBM SPSS Statistics*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Winer B.J. (1971), *Statistical principles in experimental design*. Second Edition, McGraw-Hill, New York, NY.

Winer B.J., Brown D.R., Michels K.M. (1991), *Statistical principles in experimental design,* McGraw-Hill, New York, NY.

ANEKS

Załącznik 1. Moc testów w modelu $T \times r$

Tabela 27. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości próby $n = 200$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{T\times r}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{T\times r}$	$moc_{T\times r}$
2	5	20	200	0.01	0.0100	1	4	7.709	1.667	0.1710	4	190	2.419	1.200	0.0938
2	10	10	200	0.01	0.0100	1	9	5.117	1.818	0.2267	9	180	1.932	1.100	0.0794
2	20	5	200	0.01	0.0100	1	19	4.381	1.905	0.2587	19	160	1.652	1.050	0.0688
2	50	2	200	0.01	0.0100	1	49	4.038	1.961	0.2790	49	100	1.480	1.020	0.0592
2	5	20	200	0.01	0.0625	1	4	7.709	0.889	0.1141	4	190	2.419	2.250	0.3701
2	10	10	200	0.01	0.0625	1	9	5.117	1.231	0.1687	9	180	1.932	1.625	0.3045
2	20	5	200	0.01	0.0625	1	19	4.381	1.524	0.2164	19	160	1.652	1.313	0.2183
2	50	2	200	0.01	0.0625	1	49	4.038	1.778	0.2575	49	100	1.480	1.125	0.1243
2	5	20	200	0.01	0.1600	1	4	7.709	0.476	0.0842	4	190	2.419	4.200	0.6804
2	10	10	200	0.01	0.1600	1	9	5.117	0.769	0.1234	9	180	1.932	2.600	0.6690
2	20	5	200	0.01	0.1600	1	19	4.381	1.111	0.1704	19	160	1.652	1.800	0.5619
2	50	2	200	0.01	0.1600	1	49	4.038	1.515	0.2265	49	100	1.480	1.320	0.3104
2	5	20	200	0.0625	0.0100	1	4	7.709	10.417	0.6793	4	190	2.419	1.200	0.0938
2	10	10	200	0.0625	0.0100	1	9	5.117	11.364	0.8499	9	180	1.932	1.100	0.0794
2	20	5	200	0.0625	0.0100	1	19	4.381	11.905	0.9051	19	160	1.652	1.050	0.0688
2	50	2	200	0.0625	0.0100	1	49	4.038	12.255	0.9294	49	100	1.480	1.020	0.0592
2	5	20	200	0.0625	0.0625	1	4	7.709	5.556	0.4360	4	190	2.419	2.250	0.3701
2	10	10	200	0.0625	0.0625	1	9	5.117	7.692	0.6952	9	180	1.932	1.625	0.3045
2	20	5	200	0.0625	0.0625	1	19	4.381	9.524	0.8332	19	160	1.652	1.313	0.2183
2	50	2	200	0.0625	0.0625	1	49	4.038	11.111	0.9045	49	100	1.480	1.125	0.1243
2	5	20	200	0.0625	0.1600	1	4	7.709	2.976	0.2653	4	190	2.419	4.200	0.6804
2	10	10	200	0.0625	0.1600	1	9	5.117	4.808	0.4990	9	180	1.932	2.600	0.6690
2	20	5	200	0.0625	0.1600	1	19	4.381	6.944	0.7055	19	160	1.652	1.800	0.5619
2	50	2	200	0.0625	0.1600	1	49	4.038	9.470	0.8546	49	100	1.480	1.320	0.3104
2	5	20	200	0.16	0.0100	1	4	7.709	26.667	0.9644	4	190	2.419	1.200	0.0938
2	10	10	200	0.16	0.0100	1	9	5.117	29.091	0.9974	9	180	1.932	1.100	0.0794

2	20	5	200	0.16	0.0100	1	19	4.381	30.476	0.9995	19	160	1.652	1.050	0.0688
2	50	2	200	0.16	0.0100	1	49	4.038	31.373	0.9998	49	100	1.480	1.020	0.0592
2	5	20	200	0.16	0.0625	1	4	7.709	14.222	0.8020	4	190	2.419	2.250	0.3701
2	10	10	200	0.16	0.0625	1	9	5.117	19.692	0.9755	9	180	1.932	1.625	0.3045
2	20	5	200	0.16	0.0625	1	19	4.381	24.381	0.9967	19	160	1.652	1.313	0.2183
2	50	2	200	0.16	0.0625	1	49	4.038	28.444	0.9995	49	100	1.480	1.125	0.1243
2	5	20	200	0.16	0.1600	1	4	7.709	7.619	0.5523	4	190	2.419	4.200	0.6804
2	10	10	200	0.16	0.1600	1	9	5.117	12.308	0.8762	9	180	1.932	2.600	0.6690
2	20	5	200	0.16	0.1600	1	19	4.381	17.778	0.9791	19	160	1.652	1.800	0.5619
2	50	2	200	0.16	0.1600	1	49	4.038	24.242	0.9979	49	100	1.480	1.320	0.3104

Źródło: opracowanie własne.

Tabela 28. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości próby $n = 400$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{T\times r}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{T\times r}$	$moc_{T\times r}$
2	5	40	400	0.01	0.0100	1	4	7.709	2.857	0.2568	4	390	2.395	1.400	0.1468
2	10	20	400	0.01	0.0100	1	9	5.117	3.333	0.3717	9	380	1.905	1.200	0.1171
2	20	10	400	0.01	0.0100	1	19	4.381	3.636	0.4406	19	360	1.616	1.100	0.0935
2	50	4	400	0.01	0.0100	1	49	4.038	3.846	0.4852	49	300	1.397	1.040	0.0733
2	5	40	400	0.01	0.0625	1	4	7.709	1.143	0.1327	4	390	2.395	3.500	0.6032
2	10	20	400	0.01	0.0625	1	9	5.117	1.778	0.2227	9	380	1.905	2.250	0.5737
2	20	10	400	0.01	0.0625	1	19	4.381	2.462	0.3195	19	360	1.616	1.625	0.4671
2	50	4	400	0.01	0.0625	1	49	4.038	3.200	0.4184	49	300	1.397	1.250	0.2850
2	5	40	400	0.01	0.1600	1	4	7.709	0.541	0.0888	4	390	2.395	7.400	0.8621
2	10	20	400	0.01	0.1600	1	9	5.117	0.952	0.1413	9	380	1.905	4.200	0.9049
2	20	10	400	0.01	0.1600	1	19	4.381	1.538	0.2180	19	360	1.616	2.600	0.8903
2	50	4	400	0.01	0.1600	1	49	4.038	2.439	0.3343	49	300	1.397	1.640	0.7481
2	5	40	400	0.0625	0.0100	1	4	7.709	17.857	0.8779	4	390	2.395	1.400	0.1468
2	10	20	400	0.0625	0.0100	1	9	5.117	20.833	0.9812	9	380	1.905	1.200	0.1171
2	20	10	400	0.0625	0.0100	1	19	4.381	22.727	0.9947	19	360	1.616	1.100	0.0935
2	50	4	400	0.0625	0.0100	1	49	4.038	24.038	0.9978	49	300	1.397	1.040	0.0733
2	5	40	400	0.0625	0.0625	1	4	7.709	7.143	0.5272	4	390	2.395	3.500	0.6032
2	10	20	400	0.0625	0.0625	1	9	5.117	11.111	0.8421	9	380	1.905	2.250	0.5737
2	20	10	400	0.0625	0.0625	1	19	4.381	15.385	0.9606	19	360	1.616	1.625	0.4671
2	50	4	400	0.0625	0.0625	1	49	4.038	20.000	0.9923	49	300	1.397	1.250	0.2850
2	5	40	400	0.0625	0.1600	1	4	7.709	3.378	0.2934	4	390	2.395	7.400	0.8621
2	10	20	400	0.0625	0.1600	1	9	5.117	5.952	0.5855	9	380	1.905	4.200	0.9049
2	20	10	400	0.0625	0.1600	1	19	4.381	9.615	0.8367	19	360	1.616	2.600	0.8903
2	50	4	400	0.0625	0.1600	1	49	4.038	15.244	0.9690	49	300	1.397	1.640	0.7481
2	5	40	400	0.16	0.0100	1	4	7.709	45.714	0.9979	4	390	2.395	1.400	0.1468
2	10	20	400	0.16	0.0100	1	9	5.117	53.333	1.0000	9	380	1.905	1.200	0.1171
2	20	10	400	0.16	0.0100	1	19	4.381	58.182	1.0000	19	360	1.616	1.100	0.0935
2	50	4	400	0.16	0.0100	1	49	4.038	61.538	1.0000	49	300	1.397	1.040	0.0733
2	5	40	400	0.16	0.0625	1	4	7.709	18.286	0.8848	4	390	2.395	3.500	0.6032
2	10	20	400	0.16	0.0625	1	9	5.117	28.444	0.9970	9	380	1.905	2.250	0.5737
2	20	10	400	0.16	0.0625	1	19	4.381	39.385	1.0000	19	360	1.616	1.625	0.4671
2	50	4	400	0.16	0.0625	1	49	4.038	51.200	1.0000	49	300	1.397	1.250	0.2850
2	5	40	400	0.16	0.1600	1	4	7.709	8.649	0.6030	4	390	2.395	7.400	0.8621

2	10	20	400	0.16	0.1600	1	9	5.117	15.238	0.9335	9	380	1.905	4.200	0.9049
2	20	10	400	0.16	0.1600	1	19	4.381	24.615	0.9969	19	360	1.616	2.600	0.8903
2	50	4	400	0.16	0.1600	1	49	4.038	39.024	1.0000	49	300	1.397	1.640	0.7481

Źródło: opracowanie własne.

Tabela 29. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości grupy $s = 20$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{T\times r}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{T\times r}$	$moc_{T\times r}$
2	5	20	200	0.01	0.0100	1	4	7.709	1.667	0.1710	4	190	2.419	1.200	0.0938
2	10	20	400	0.01	0.0100	1	9	5.117	3.333	0.3717	9	380	1.905	1.200	0.1171
2	20	20	800	0.01	0.0100	1	19	4.381	6.667	0.6879	19	760	1.600	1.200	0.1540
2	50	20	2000	0.01	0.0100	1	49	4.038	16.667	0.9794	49	1900	1.361	1.200	0.2449
2	5	20	200	0.01	0.0625	1	4	7.709	0.889	0.1141	4	190	2.419	2.250	0.3701
2	10	20	400	0.01	0.0625	1	9	5.117	1.778	0.2227	9	380	1.905	2.250	0.5737
2	20	20	800	0.01	0.0625	1	19	4.381	3.556	0.4327	19	760	1.600	2.250	0.8096
2	50	20	2000	0.01	0.0625	1	49	4.038	8.889	0.8321	49	1900	1.361	2.250	0.9864
2	5	20	200	0.01	0.1600	1	4	7.709	0.476	0.0842	4	190	2.419	4.200	0.6804
2	10	20	400	0.01	0.1600	1	9	5.117	0.952	0.1413	9	380	1.905	4.200	0.9049
2	20	20	800	0.01	0.1600	1	19	4.381	1.905	0.2587	19	760	1.600	4.200	0.9925
2	50	20	2000	0.01	0.1600	1	49	4.038	4.762	0.5712	49	1900	1.361	4.200	1.0000
2	5	20	200	0.0625	0.0100	1	4	7.709	10.417	0.6793	4	190	2.419	1.200	0.0938
2	10	20	400	0.0625	0.0100	1	9	5.117	20.833	0.9812	9	380	1.905	1.200	0.1171
2	20	20	800	0.0625	0.0100	1	19	4.381	41.667	1.0000	19	760	1.600	1.200	0.1540
2	50	20	2000	0.0625	0.0100	1	49	4.038	104.167	1.0000	49	1900	1.361	1.200	0.2449
2	5	20	200	0.0625	0.0625	1	4	7.709	5.556	0.4360	4	190	2.419	2.250	0.3701
2	10	20	400	0.0625	0.0625	1	9	5.117	11.111	0.8421	9	380	1.905	2.250	0.5737
2	20	20	800	0.0625	0.0625	1	19	4.381	22.222	0.9939	19	760	1.600	2.250	0.8096
2	50	20	2000	0.0625	0.0625	1	49	4.038	55.556	1.0000	49	1900	1.361	2.250	0.9864
2	5	20	200	0.0625	0.1600	1	4	7.709	2.976	0.2653	4	190	2.419	4.200	0.6804
2	10	20	400	0.0625	0.1600	1	9	5.117	5.952	0.5855	9	380	1.905	4.200	0.9049
2	20	20	800	0.0625	0.1600	1	19	4.381	11.905	0.9051	19	760	1.600	4.200	0.9925
2	50	20	2000	0.0625	0.1600	1	49	4.038	29.762	0.9996	49	1900	1.361	4.200	1.0000
2	5	20	200	0.16	0.0100	1	4	7.709	26.667	0.9644	4	190	2.419	1.200	0.0938
2	10	20	400	0.16	0.0100	1	9	5.117	53.333	1.0000	9	380	1.905	1.200	0.1171
2	20	20	800	0.16	0.0100	1	19	4.381	106.667	1.0000	19	760	1.600	1.200	0.1540
2	50	20	2000	0.16	0.0100	1	49	4.038	266.667	1.0000	49	1900	1.361	1.200	0.2449
2	5	20	200	0.16	0.0625	1	4	7.709	14.222	0.8020	4	190	2.419	2.250	0.3701
2	10	20	400	0.16	0.0625	1	9	5.117	28.444	0.9970	9	380	1.905	2.250	0.5737
2	20	20	800	0.16	0.0625	1	19	4.381	56.889	1.0000	19	760	1.600	2.250	0.8096
2	50	20	2000	0.16	0.0625	1	49	4.038	142.222	1.0000	49	1900	1.361	2.250	0.9864
2	5	20	200	0.16	0.1600	1	4	7.709	7.619	0.5523	4	190	2.419	4.200	0.6804

2	10	20	400	0.16	0.1600	1	9	5.117	15.238	0.9335	9	380	1.905	4.200	0.9049
2	20	20	800	0.16	0.1600	1	19	4.381	30.476	0.9995	19	760	1.600	4.200	0.9925
2	50	20	2000	0.16	0.1600	1	49	4.038	76.190	1.0000	49	1900	1.361	4.200	1.0000

Źródło: opracowanie własne.

Załącznik 2. Moc testów w modelu $r(T)$

Tabela 30. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości próby $n = 200$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{r(T)}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{r(T)}$	$moc_{r(T)}$
2	5	20	200	0.01	0.0100	1	8	5.318	1.667	0.2071	8	190	1.987	1.200	0.1116
2	10	10	200	0.01	0.0100	1	18	4.414	1.818	0.2479	18	180	1.661	1.100	0.0905
2	20	5	200	0.01	0.0100	1	38	4.098	1.905	0.2699	38	160	1.480	1.050	0.0749
2	50	2	200	0.01	0.0100	1	98	3.938	1.961	0.2836	98	100	1.394	1.020	0.0609
2	5	20	200	0.01	0.0625	1	8	5.318	0.889	0.1327	8	190	1.987	2.250	0.5316
2	10	10	200	0.01	0.0625	1	18	4.414	1.231	0.1830	18	180	1.661	1.625	0.4368
2	20	5	200	0.01	0.0625	1	38	4.098	1.524	0.2254	38	160	1.480	1.313	0.2988
2	50	2	200	0.01	0.0625	1	98	3.938	1.778	0.2617	98	100	1.394	1.125	0.1440
2	5	20	200	0.01	0.1600	1	8	5.318	0.476	0.0937	8	190	1.987	4.200	0.8741
2	10	10	200	0.01	0.1600	1	18	4.414	0.769	0.1321	18	180	1.661	2.600	0.8653
2	20	5	200	0.01	0.1600	1	38	4.098	1.111	0.1770	38	160	1.480	1.800	0.7569
2	50	2	200	0.01	0.1600	1	98	3.938	1.515	0.2301	98	100	1.394	1.320	0.3936
2	5	20	200	0.0625	0.0100	1	8	5.318	10.417	0.8063	8	190	1.987	1.200	0.1116
2	10	10	200	0.0625	0.0100	1	18	4.414	11.364	0.8900	18	180	1.661	1.100	0.0905
2	20	5	200	0.0625	0.0100	1	38	4.098	11.905	0.9195	38	160	1.480	1.050	0.0749
2	50	2	200	0.0625	0.0100	1	98	3.938	12.255	0.9340	98	100	1.394	1.020	0.0609
2	5	20	200	0.0625	0.0625	1	8	5.318	5.556	0.5443	8	190	1.987	2.250	0.5316
2	10	10	200	0.0625	0.0625	1	18	4.414	7.692	0.7464	18	180	1.661	1.625	0.4368
2	20	5	200	0.0625	0.0625	1	38	4.098	9.524	0.8525	38	160	1.480	1.313	0.2988
2	50	2	200	0.0625	0.0625	1	98	3.938	11.111	0.9100	98	100	1.394	1.125	0.1440
2	5	20	200	0.0625	0.1600	1	8	5.318	2.976	0.3304	8	190	1.987	4.200	0.8741
2	10	10	200	0.0625	0.1600	1	18	4.414	4.808	0.5458	18	180	1.661	2.600	0.8653
2	20	5	200	0.0625	0.1600	1	38	4.098	6.944	0.7285	38	160	1.480	1.800	0.7569
2	50	2	200	0.0625	0.1600	1	98	3.938	9.470	0.8615	98	100	1.394	1.320	0.3936
2	5	20	200	0.16	0.0100	1	8	5.318	26.667	0.9941	8	190	1.987	1.200	0.1116
2	10	10	200	0.16	0.0100	1	18	4.414	29.091	0.9991	18	180	1.661	1.100	0.0905
2	20	5	200	0.16	0.0100	1	38	4.098	30.476	0.9997	38	160	1.480	1.050	0.0749
2	50	2	200	0.16	0.0100	1	98	3.938	31.373	0.9998	98	100	1.394	1.020	0.0609

2	5	20	200	0.16	0.0625	1	8	5.318	14.222	0.9086	8	190	1.987	2.250	0.5316
2	10	10	200	0.16	0.0625	1	18	4.414	19.692	0.9871	18	180	1.661	1.625	0.4368
2	20	5	200	0.16	0.0625	1	38	4.098	24.381	0.9978	38	160	1.480	1.313	0.2988
2	50	2	200	0.16	0.0625	1	98	3.938	28.444	0.9996	98	100	1.394	1.125	0.1440
2	5	20	200	0.16	0.1600	1	8	5.318	7.619	0.6776	8	190	1.987	4.200	0.8741
2	10	10	200	0.16	0.1600	1	18	4.414	12.308	0.9124	18	180	1.661	2.600	0.8653
2	20	5	200	0.16	0.1600	1	38	4.098	17.778	0.9841	38	160	1.480	1.800	0.7569
2	50	2	200	0.16	0.1600	1	98	3.938	24.242	0.9982	98	100	1.394	1.320	0.3936

Źródło: opracowanie własne.

Tabela 31. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości próby $n = 400$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{r(T)}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{r(T)}$	$moc_{r(T)}$
2	5	40	400	0.01	0.0100	1	8	5.318	2.857	0.3195	8	390	1.962	1.400	0.1940
2	10	20	400	0.01	0.0100	1	18	4.414	3.333	0.4085	18	380	1.631	1.200	0.1484
2	20	10	400	0.01	0.0100	1	38	4.098	3.636	0.4597	38	360	1.439	1.100	0.1120
2	50	4	400	0.01	0.0100	1	98	3.938	3.846	0.4929	98	300	1.298	1.040	0.0811
2	5	40	400	0.01	0.0625	1	8	5.318	1.143	0.1569	8	390	1.962	3.500	0.8101
2	10	20	400	0.01	0.0625	1	18	4.414	1.778	0.2435	18	380	1.631	2.250	0.7857
2	20	10	400	0.01	0.0625	1	38	4.098	2.462	0.3336	38	360	1.439	1.625	0.6667
2	50	4	400	0.01	0.0625	1	98	3.938	3.200	0.4253	98	300	1.298	1.250	0.3984
2	5	40	400	0.01	0.1600	1	8	5.318	0.541	0.0998	8	390	1.962	7.400	0.9767
2	10	20	400	0.01	0.1600	1	18	4.414	0.952	0.1522	18	380	1.631	4.200	0.9895
2	20	10	400	0.01	0.1600	1	38	4.098	1.538	0.2271	38	360	1.439	2.600	0.9860
2	50	4	400	0.01	0.1600	1	98	3.938	2.439	0.3398	98	300	1.298	1.640	0.9135
2	5	40	400	0.0625	0.0100	1	8	5.318	17.857	0.9574	8	390	1.962	1.400	0.1940
2	10	20	400	0.0625	0.0100	1	18	4.414	20.833	0.9906	18	380	1.631	1.200	0.1484
2	20	10	400	0.0625	0.0100	1	38	4.098	22.727	0.9964	38	360	1.439	1.100	0.1120
2	50	4	400	0.0625	0.0100	1	98	3.938	24.038	0.9981	98	300	1.298	1.040	0.0811
2	5	40	400	0.0625	0.0625	1	8	5.318	7.143	0.6499	8	390	1.962	3.500	0.8101
2	10	20	400	0.0625	0.0625	1	18	4.414	11.111	0.8832	18	380	1.631	2.250	0.7857
2	20	10	400	0.0625	0.0625	1	38	4.098	15.385	0.9687	38	360	1.439	1.625	0.6667
2	50	4	400	0.0625	0.0625	1	98	3.938	20.000	0.9932	98	300	1.298	1.250	0.3984
2	5	40	400	0.0625	0.1600	1	8	5.318	3.378	0.3668	8	390	1.962	7.400	0.9767
2	10	20	400	0.0625	0.1600	1	18	4.414	5.952	0.6362	18	380	1.631	4.200	0.9895
2	20	10	400	0.0625	0.1600	1	38	4.098	9.615	0.8558	38	360	1.439	2.600	0.9860
2	50	4	400	0.0625	0.1600	1	98	3.938	15.244	0.9717	98	300	1.298	1.640	0.9135
2	5	40	400	0.16	0.0100	1	8	5.318	45.714	0.9999	8	390	1.962	1.400	0.1940
2	10	20	400	0.16	0.0100	1	18	4.414	53.333	1.0000	18	380	1.631	1.200	0.1484
2	20	10	400	0.16	0.0100	1	38	4.098	58.182	1.0000	38	360	1.439	1.100	0.1120
2	50	4	400	0.16	0.0100	1	98	3.938	61.538	1.0000	98	300	1.298	1.040	0.0811
2	5	40	400	0.16	0.0625	1	8	5.318	18.286	0.9611	8	390	1.962	3.500	0.8101
2	10	20	400	0.16	0.0625	1	18	4.414	28.444	0.9989	18	380	1.631	2.250	0.7857
2	20	10	400	0.16	0.0625	1	38	4.098	39.385	1.0000	38	360	1.439	1.625	0.6667
2	50	4	400	0.16	0.0625	1	98	3.938	51.200	1.0000	98	300	1.298	1.250	0.3984

2	5	40	400	0.16	0.1600		1	8	5.318	8.649	0.7314		8	390	1.962	7.400	0.9767
2	10	20	400	0.16	0.1600		1	18	4.414	15.238	0.9580		18	380	1.631	4.200	0.9895
2	20	10	400	0.16	0.1600		1	38	4.098	24.615	0.9980		38	360	1.439	2.600	0.9860
2	50	4	400	0.16	0.1600		1	98	3.938	39.024	1.0000		98	300	1.298	1.640	0.9135

Źródło: opracowanie własne.

Tabela 32. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości grupy $s = 20$, $alfa = 0.05$

t	r	s	n	f_T^2	$f_{r(T)}^2$	df_1	df_2	F_{kryt}	λ_T	moc_T	df_1	df_2	F_{kryt}	$\kappa_{r(T)}$	$moc_{r(T)}$
2	5	20	200	0.01	0.0100	1	8	5.318	1.667	0.2071	8	190	1.987	1.200	0.1116
2	10	20	400	0.01	0.0100	1	18	4.414	3.333	0.4085	18	380	1.631	1.200	0.1484
2	20	20	800	0.01	0.0100	1	38	4.098	6.667	0.7110	38	760	1.421	1.200	0.2099
2	50	20	2000	0.01	0.0100	1	98	3.938	16.667	0.9813	98	1900	1.254	1.200	0.3635
2	5	20	200	0.01	0.0625	1	8	5.318	0.889	0.1327	8	190	1.987	2.250	0.5316
2	10	20	400	0.01	0.0625	1	18	4.414	1.778	0.2435	18	380	1.631	2.250	0.7857
2	20	20	800	0.01	0.0625	1	38	4.098	3.556	0.4515	38	760	1.421	2.250	0.9602
2	50	20	2000	0.01	0.0625	1	98	3.938	8.889	0.8394	98	1900	1.254	2.250	0.9999
2	5	20	200	0.01	0.1600	1	8	5.318	0.476	0.0937	8	190	1.987	4.200	0.8741
2	10	20	400	0.01	0.1600	1	18	4.414	0.952	0.1522	18	380	1.631	4.200	0.9895
2	20	20	800	0.01	0.1600	1	38	4.098	1.905	0.2699	38	760	1.421	4.200	0.9999
2	50	20	2000	0.01	0.1600	1	98	3.938	4.762	0.5796	98	1900	1.254	4.200	1.0000
2	5	20	200	0.0625	0.0100	1	8	5.318	10.417	0.8063	8	190	1.987	1.200	0.1116
2	10	20	400	0.0625	0.0100	1	18	4.414	20.833	0.9906	18	380	1.631	1.200	0.1484
2	20	20	800	0.0625	0.0100	1	38	4.098	41.667	1.0000	38	760	1.421	1.200	0.2099
2	50	20	2000	0.0625	0.0100	1	98	3.938	104.167	1.0000	98	1900	1.254	1.200	0.3635
2	5	20	200	0.0625	0.0625	1	8	5.318	5.556	0.5443	8	190	1.987	2.250	0.5316
2	10	20	400	0.0625	0.0625	1	18	4.414	11.111	0.8832	18	380	1.631	2.250	0.7857
2	20	20	800	0.0625	0.0625	1	38	4.098	22.222	0.9958	38	760	1.421	2.250	0.9602
2	50	20	2000	0.0625	0.0625	1	98	3.938	55.556	1.0000	98	1900	1.254	2.250	0.9999
2	5	20	200	0.0625	0.1600	1	8	5.318	2.976	0.3304	8	190	1.987	4.200	0.8741
2	10	20	400	0.0625	0.1600	1	18	4.414	5.952	0.6362	18	380	1.631	4.200	0.9895
2	20	20	800	0.0625	0.1600	1	38	4.098	11.905	0.9195	38	760	1.421	4.200	0.9999
2	50	20	2000	0.0625	0.1600	1	98	3.938	29.762	0.9997	98	1900	1.254	4.200	1.0000
2	5	20	200	0.16	0.0100	1	8	5.318	26.667	0.9941	8	190	1.987	1.200	0.1116
2	10	20	400	0.16	0.0100	1	18	4.414	53.333	1.0000	18	380	1.631	1.200	0.1484
2	20	20	800	0.16	0.0100	1	38	4.098	106.667	1.0000	38	760	1.421	1.200	0.2099
2	50	20	2000	0.16	0.0100	1	98	3.938	266.667	1.0000	98	1900	1.254	1.200	0.3635
2	5	20	200	0.16	0.0625	1	8	5.318	14.222	0.9086	8	190	1.987	2.250	0.5316
2	10	20	400	0.16	0.0625	1	18	4.414	28.444	0.9989	18	380	1.631	2.250	0.7857
2	20	20	800	0.16	0.0625	1	38	4.098	56.889	1.0000	38	760	1.421	2.250	0.9602
2	50	20	2000	0.16	0.0625	1	98	3.938	142.222	1.0000	98	1900	1.254	2.250	0.9999
2	5	20	200	0.16	0.1600	1	8	5.318	7.619	0.6776	8	190	1.987	4.200	0.8741

2	10	20	400	0.16	0.1600	1	18	4.414	15.238	0.9580	18	380	1.631	4.200	0.9895
2	20	20	800	0.16	0.1600	1	38	4.098	30.476	0.9997	38	760	1.421	4.200	0.9999
2	50	20	2000	0.16	0.1600	1	98	3.938	76.190	1.0000	98	1900	1.254	4.200	1.0000

Źródło: opracowanie własne.

Załącznik 3. Dane do przykładu

Tabela 33. Dane wykorzystane w analizie

T	***r*** **(w strukturze krzyżowej)**	***r*** **(w strukturze zagnieżdżonej)**	***s***	***indeks***
1	1	1	1	14
1	1	1	2	13
1	1	1	3	17
1	1	1	4	15
1	1	1	5	14
1	2	2	6	26
1	2	2	7	15
1	2	2	8	21
1	2	2	9	20
1	2	2	10	26
1	3	3	11	15
1	3	3	12	21
1	3	3	13	18
1	3	3	14	16
1	3	3	15	22
2	1	4	16	19
2	1	4	17	18
2	1	4	18	15
2	1	4	19	18
2	1	4	20	19
2	2	5	21	14
2	2	5	22	16
2	2	5	23	15
2	2	5	24	15
2	2	5	25	12
2	3	6	26	10
2	3	6	27	13
2	3	6	28	13
2	3	6	29	15
2	3	6	30	10

Źródło: opracowanie własne.

Załącznik 4. Komendy do Edytora Poleceń w IBM® SPSS® Statistics

```
****Tabela 18. i tabela 19.
GLM indeks BY T r s
 /RANDOM=r s
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
 /EMMEANS=TABLES (T)
 /EMMEANS=TABLES (r)
 /EMMEANS=TABLES (T*r)
 /DESIGN=T r T*r s(T*r)
 /TEST T VS T*r
 /TEST r VS s(T*r)
 /TEST T*r VS s(T*r).

****Tabela 20.
GLM indeks BY T r
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
 /DESIGN=T r T*r.

****Tabela 21.
GLM indeks BY T
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
 /EMMEANS=TABLES (T)
 /DESIGN=T.

****Tabela 22. i tabela 23.
GLM indeks BY T r s
 /RANDOM=r s
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
```

```
 /EMMEANS=TABLES (T)
 /EMMEANS=TABLES (r)
 /DESIGN=T r(T) s(r(T))
 /TEST T VS r(T)
 /TEST r(T) VS s(r(T)).

****Tabela 24.
GLM indeks BY T r
 /METHOD=SSTYPE(3)
 /INTERCEPT=INCLUDE
 /CRITERIA=ALPHA(0.05)
 /DESIGN=T r(T).

****Tabela 25.
SORT CASES BY T.
SPLIT FILE LAYERED BY T.
GLM indeks BY r
/DESIGN=r.

SPLIT FILE OFF.
```

Spis rysunków

Rysunek 1. Schemat tabeli danych dla planu 1a 30
Rysunek 2. Schemat tabeli danych dla planu 2a 31
Rysunek 3. Schemat tabeli danych dla planu 1b 31
Rysunek 4. Schemat tabeli danych dla planu 2b 32
Rysunek 5. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 1. 87
Rysunek 6. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 2. 87
Rysunek 7. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 3. 88
Rysunek 8. Wynik losowania trzech elementów z każdej z dwóch populacji – przypadek 4. 88
Rysunek 9. Wynik losowania trzech elementów z jednej populacji. 90

Spis tabel

Tabela 1. Oczekiwane średnie kwadraty dla schematu z efektem stałym T 69
Tabela 2. Oczekiwane średnie kwadraty dla schematu efektami stałymi w strukturze krzyżowej $T \times R$ 70
Tabela 3. Realizacja kroku III a dla schematu 2b 80
Tabela 4. Realizacja kroku III b dla schematu 2b 81
Tabela 5. Realizacja kroku III c dla schematu 2b 82
Tabela 6. Realizacja kroku III d dla schematu 2b. Oczekiwane średnie kwadraty dla schematu z efektem stałym T oraz efektem losowym r w strukturze krzyżowej $T \times r$ 82
Tabela 7. Realizacja kroku III a dla schematu 2a 83
Tabela 8. Realizacja kroku III b dla schematu 2a 83
Tabela 9. Realizacja kroku III c dla schematu 2a 84
Tabela 10. Realizacja kroku III d dla schematu 2a. Oczekiwane średnie kwadraty dla schematu z efektem stałym T oraz efektem losowym r w strukturze zagnieżdżonej $r(T)$ 84
Tabela 11. Źródła zmienności kształtujące średnie kwadraty dla poszczególnych efektów w schemacie 2b (modelu $T \times r$) 85
Tabela 12. Źródła zmienności kształtujące średnie kwadraty dla poszczególnych efektów w schemacie 2a (modelu $r(T)$) 85
Tabela 13. Oczekiwane średnie kwadraty dla schematu $s \times r(T)$ 90
Tabela 14. Oczekiwane średnie kwadraty dla schematu z efektem losowym t 100
Tabela 15. Oczekiwane średnie kwadraty dla schematu z efektami losowymi $t \times r$ 103
Tabela 16. Oczekiwane średnie kwadraty dla schematu z efektami losowymi $r(t)$ 109
Tabela 17. Oczekiwane średnie kwadraty dla schematu z efektami stałymi $R(T)$ 110
Tabela 18. Średnie grupowe dla modelu $T \times r$ 120
Tabela 19. Wyniki analizy wariancji dla modelu $T \times r$ (czynnik r losowy) 120
Tabela 20. Wyniki analizy wariancji dla modelu $T \times R$ (czynnik R stały) . 122
Tabela 21. Wyniki analizy wariancji dla modelu T (czynnik r pominięty) 123
Tabela 22. Średnie grupowe dla modelu $r(T)$ 124
Tabela 23. Wyniki analizy wariancji dla modelu $r(T)$ (czynnik r losowy) 125
Tabela 24 Wyniki analizy wariancji dla modelu $R(T)$ (czynnik R stały) ... 126

Tabela 25. Sumy kwadratów odchyleń w modelu $r(T)$ wraz z dekompozycją zmienności dla czynnika $r(T)$ 127
Tabela 26. Sumy kwadratów i stopnie swobody dla poszczególnych źródeł w modelach ze strukturą krzyżową i zagnieżdżoną 128
Tabela 27. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości próby $n = 200$, $alfa = 0.05$ 143
Tabela 28. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości próby $n = 400$, $alfa = 0.05$ 145
Tabela 29. Moc dla testu efektu T oraz testu efektu $T \times r$ przy stałej wielkości grupy $s = 20$, $alfa = 0.05$ 147
Tabela 30. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości próby $n = 200$, $alfa = 0.05$ 149
Tabela 31. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości próby $n = 400$, $alfa = 0.05$ 151
Tabela 32. Moc dla testu efektu T oraz testu efektu $r(T)$ przy stałej wielkości grupy $s = 20$, $alfa = 0.05$ 153
Tabela 33. Dane wykorzystane w analizie 155

Spis wykresów

Wykres 1. Efekt interakcji $T \times r$ 121
Wykres 2. Efekt czynnika $r(T)$ 126

Katarzyna Grzeszkiewicz-Radulska – adiunkt w Katedrze Metod i Technik Badań Społecznych Instytutu Socjologii Uniwersytetu Łódzkiego. Autorka książki *Respondenci niedostępni w badaniach sondażowych* (2009), a także współautorka publikacji *Analizy weryfikacyjne – przeszłe i obecne doświadczenia badawcze* (2017) oraz *Analiza statystyczna z IBM SPSS Statistics* (2020).

Głównymi adresatami książki są badacze korzystający z metody eksperymentu, w szczególności w naukach społecznych. Czytelnicy dowiedzą się, dlaczego rozwiązanie wielu typowych dla tych nauk problemów badawczych powinno nastąpić poprzez włączenie czynników losowych do planu eksperymentalnego oraz dlaczego zaniechanie tej czynności może prowadzić do ustaleń o niskiej trafności, a nawet do ustaleń fałszywie pozytywnych. Autorka szeroko prezentuje stronę analityczną zagadnienia – pokazuje, w jaki sposób przeprowadzić analizę wariancji (ANOVA), gdy w modelu występują zarówno czynniki stałe, jak i losowe. Pokazuje tym samym, jak uogólniać wnioski na kilka populacji jednocześnie – nie tylko na populację jednostek, osób czy respondentów, lecz także na inne zbiorowości, którymi równolegle mogą być populacja reklam, słów, ankieterów itd.

Publikacja, ze względu na prezentację trudnych treści w sposób przystępny i zrozumiały, [...] otwiera możliwości do świadomego, poprawnego stosowania analizy wariancji przy różnych schematach (w tym z czynnikami losowymi).

Z recenzji dr hab. Jolanty Perek-Białas, prof. UJ

Książka jest ciekawą pozycją [...] z zakresu metodologii badań społecznych i stanowi ona uzupełnienie wcześniejszych opracowań dotyczących podobnych kwestii. [...] zawiera wiele przykładów wykorzystania analizy wariancji z czynnikami losowymi do rozwiązywania ważkich problemów w socjologii (i nie tylko).

Z recenzji dr. hab. Piotra Jabkowskiego, prof. UAM

ISBN 978-83-8331-176-0

WYDAWNICTWO UNIWERSYTETU ŁÓDZKIEGO

wydawnictwo.uni.lodz.pl
ksiegarnia@uni.lodz.pl
(42) 665 58 63

Książka dostępna również jako e-book