*Psychologia Społeczna*
*2010 tom 5 2–3 (14) 234–259*

ISSN 1896-1800

# Schemat powtarzanych pomiarów w ujęciu analizy wielopoziomowej – hierarchiczny model liniowy jako alternatywa dla analizy wariancji z powtarzanym pomiarem

Piotr Zieliński

*Wojskowy Instytut Medycyny Lotniczej*

Plany z powtarzanymi pomiarami to często stosowany schemat badawczy, mający w niektórych sytuacjach znaczną przewagę nad badaniami w planie grup niezależnych. Najczęściej stosowaną metodą analizy danych z powtarzanych pomiarów jest jednozmiennowa analiza wariancji (ANOVA) w tzw. modelu efektów mieszanych. Analiza ta oparta jest jednak na ścisłych założeniach, które nie zawsze mogą być spełnione. Niniejszy tekst prezentuje możliwość analizy powtarzanych pomiarów w oparciu o hierarchiczny model liniowy – metodę cechującą się bardziej liberalnymi założeniami i większą elastycznością w modelowaniu danych. Dwa opisane w tekście przykłady pokazują, że w prostych planach badawczych wyniki uzyskane w ramach analizy wariancji i hierarchicznego modelu liniowego są niemal identyczne, w sytuacjach zaś bardziej złożonych hierarchiczny model liniowy umożliwia przeprowadzenie trafniejszej analizy niż ANOVA.

*Słowa kluczowe:* plany z powtarzanym pomiarem; hierarchiczny model liniowy; analiza wariancji dla powtarzanych pomiarów

## Schemat z powtarzanym pomiarem jako przykład modelu mieszanego

### Czym są powtarzane pomiary?

O powtarzanych pomiarach zwykliśmy mówić, gdy ta sama zmienna charakteryzująca dany obiekt bądź osobę jest mierzona kilkakrotnie (i, niekiedy, w odmiennych warunkach). Jeśli głównym predyktorem jest czas, a badacza interesują przede wszystkim zmiany w natężeniu zmiennej zależnej w okresie objętym badaniem, możemy mówić o tzw. badaniach podłużnych. Sformułowanie „plan badawczy z powtarzanymi pomiarami” jest natomiast często stosowane w kontekście eksperymentów, w których ta sama osoba badana jest kolejno w różnych warunkach eksperymentalnych.

Do podstawowych zalet powtarzanych pomiarów można zaliczyć możliwość wyłączenia różnic indywidualnych z wariancji błędu, dzięki czemu testowanie związków między interesującymi badacza zmiennymi może być efektywniejsze niż w analogicznych badaniach realizowanych w planie grup niezależnych. Równocześnie wprowadzenie powtarzanych pomiarów jest bardziej ekonomiczne, gdyż od tego samego uczestnika badania uzyskujemy wyniki z kilku sytuacji pomiarowych (np. zarówno z warunków eksperymentalnych, jak i kontrolnych), co zmniejsza liczbę osób niezbędną do uzyskania zadowalającej mocy testów statystycznych. Schemat powtarzanych pomiarów jest też jedyną metodą pozwalającą uchwycić indywidualne wzorce zmiany w obrębie zmiennej zależnej u poszczególnych jednostek. Dość przystępne wprowadzenie do planów z powtarzanymi pomiarami, omówienie ich wad oraz zalet można znaleźć w dostępnej w języku polskim literaturze np. w podręczniku *Metody badawcze w psychologii* (Shaughnessy, Zechmeister i Zechmeister, 2002).

Plany z powtarzanymi pomiarami mogą mieć charakter planów zrównoważonych (*balanced design*), w których

Piotr Zieliński, Zakład Psychologii Lotniczej, Wojskowy Instytut Medycyny Lotniczej, ul. Krasińskiego 54, 01–755 Warszawa, e-mail: pzielins@wiml.waw.pl



Copyright 2010 Psychologia Społeczna

każda z osób jest poddawana pomiarowi tyle samo razy i z zapewnieniem tych samych warunków, lub niezrównoważonych (*unbalanced design*), gdy u poszczególnych osób liczba pomiarów może być różna, pomiary mogą być dokonywane w odmiennych warunkach bądź w innych odstępach czasu (jest to spotykane szczególnie często w długo trwających badaniach podłużnych, gdy badacz nie zawsze ma kontrolę nad tym, kiedy osoba uczestnicząca w badaniu może zostać poddana pomiarowi). W długo trwających badaniach częstym problemem są też brakujące obserwacje, związane z niemożliwością dotarcia do badanych we wszystkich planowanych momentach. Może to sprawić, że nawet w planie badawczym, który z założenia był planem zrównoważonym, w efekcie wielu braków danych należy traktować wyniki jako niezrównoważone. Zależnie od struktury danych, odmienne metody statystyczne mogą okazać się właściwe dla analizy wyników. Jedną z klasycznych metod dla planów zrównoważonych jest jednozmiennowa analiza wariancji dla powtarzanych pomiarów (*univariate repeated measures*, ANOVA). W niniejszym artykule zostanie ona zestawiona z metodą opartą na analizie wielopoziomowej (*multilevel analysis*), coraz częściej polecaną przez współczesne podręczniki (np. Davis, 2002; Hedeker i Gibbons, 2006) jako bardziej uniwersalną i dającą znacznie większe możliwości modelowania struktury uzyskanych wyników.

### Analiza wariancji dla powtarzanych pomiarów – podstawowe informacje

W dostępnych obecnie w języku polskim podręcznikach (np. Bedyńska i Brzezicka, 2007; Brzeziński, 2008; Ferguson i Takane, 1997), jednozmiennowa analiza wariancji (ANOVA) jest jedną z najczęściej opisywanych metod służących do analizy danych z powtarzanymi pomiarami. Metoda ta nadaje się do analizy danych zrównoważonych (*balanced data*), jest więc szczególnie przydatna np. w schematach eksperymentalnych ze stałą liczbą powtarzanych pomiarów, w których brakujące obserwacje stanowią niewielki odsetek uzyskiwanych wyników.

Ponieważ od każdej z osób badanych uzyskujemy kilka wyników (w liczbie równej liczbie sytuacji pomiarowych), w analizie wariancji dla powtarzanych pomiarów staramy się kontrolować dwa podstawowe źródła zmienności: jedno związane z indywidualnym zróżnicowaniem osób badanych i drugie związane ze zróżnicowaniem sytuacji pomiarowych. Sytuacje pomiarowe traktujemy jako tzw. czynnik stały (*fixed factor*), co oznacza, że zbiór tych sytuacji uznajemy za stały i kompletny zbiór interesujących badacza warunków. Osoby badane mają status czynnika losowego (*random factor*), co oznacza, że zbiór osób badanych traktujemy jedynie jako zbiór losowo wybranych osób z interesującej badacza populacji. Ponieważ w tej samej analizie mamy zarówno czynnik stały, jak i czynnik losowy, cały model ma charakter tzw. modelu mieszanego (*mixed model*). Można powiedzieć, że analiza wariancji z powtarzanym pomiarem w obrębie jednego czynnika (sytuacje pomiarowe) jest po prostu dwuczynnikową analizą wariancji (pomiar *x* osoba) w modelu mieszanym z jedną obserwacją na klatkę tabeli.

Podstawowym krokiem w tej metodzie jest podział wariancji całkowitej na wariancję międzyosobową (związaną ze zróżnicowaniem wyników u osób badanych) i wewnątrzosobową (związaną ze zróżnicowaniem w obrębie wyników tej samej osoby). Jeżeli osoby badane są traktowane jako losowa próbka z jednorodnej populacji, tj. nie są wprowadzone żadne zmienne międzygrupowe, zróżnicowanie międzyosobowe z reguły nie jest dla badacza interesujące. Wariancja wewnątrzosobowa jest zaś dzielona na część związaną ze zróżnicowaniem sytuacji pomiarowych (poszczególne warunki eksperymentalne) oraz na część, której samo zróżnicowanie sytuacji pomiarowych nie tłumaczy, czyli wariancję resztową (błąd pomiaru). Dzięki takiemu postępowaniu różnice indywidualne są wyłączone z tego składnika błędu, a generowane następnie testy $F$ dla ustalenia istotności analizowanych efektów pozwalają w łatwiejszy sposób wychwycić wewnątrzosobowy wpływ zmiennych niezależnych. W modelu takim jest też oczywiście możliwe włączenie i testowanie dodatkowych zmiennych międzyosobowych i interakcji między- i wewnątrzosobowych zmiennych.

Założenia leżące u podstaw analizy wariancji z powtarzanym pomiarem dotyczą m.in. normalności rozkładu błędów oraz niezależności pomiarów u różnych osób badanych. Specyficznym założeniem jest założenie sferyczności – oznacza ono, że po obliczeniu różnic dla wszystkich par wyników z powtarzanych obserwacji, wariancje tych różnic będą do siebie zbliżone. Spełnienie tego warunku jest niezbędne do przeprowadzenia nieobciążonych testów istotności – będzie on szerzej omówiony w dalszej części niniejszego artykułu.

Bardziej szczegółowo o analizie wariancji dla powtarzanych pomiarów można przeczytać np. w podręcznikach Brzezińskiego i Stachowskiego (1984) lub Fergusona i Takane (1997).

### Hierarchiczny model liniowy jako alternatywny sposób ujęcia schematu z powtarzanym pomiarem

W ostatnich latach, dzięki coraz szerszej dostępności oprogramowania statystycznego, ważne miejsce wśród metod analizy danych zyskują tzw. analizy wielopoziomowe (*multilevel analysis*). Zależnie od tego, jakie aspekty tych metod mają być uwypuklone, inne często stosowane określenia to liniowe modele mieszane (*linear mixed effects models*) czy hierarchiczne modele liniowe (*hierarchical linear models*). Jedną z cech tego podejścia jest uwzględnianie hierarchicznego układu danych, w którym obserwacje są pogrupowane zgodnie z jakąś nadrzędną strukturą. W analizie dopuszczamy możliwość, że podobieństwo wyników w obrębie takiej grupy jest większe niż podobieństwo wyników między grupami, co sprawia, że nie można takich danych traktować jako całkowicie niezależnych. Standardowym przykładem (np. Goldstein, 1999; Hox, 2002) mogą być wyniki uczniów (najniższy poziom modelu), którzy są zagnieżdżeni w klasach (poziom drugi), które z kolei zagnieżdżone są w szkołach (poziom trzeci) itd. Uczniowie tej samej klasy wykazują większe podobieństwo między sobą niż uczniowie różnych klas, analogicznie – klasy w tej samej szkole wykazują większe podobieństwo niż klasy z różnych, często odległych geograficznie szkół. Aby uwzględnić to zróżnicowanie, dla każdego z poziomów struktury (klas, szkół) konstruowane są oddzielne składniki błędu, dzięki czemu wysokość i nachylenie linii regresji nie muszą być w poszczególnych grupach identyczne. Na każdym z tych poziomów można też wprowadzać specyficzne zmienne wyjaśniające (np. doświadczenie nauczycieli jako zmienna poziomu drugiego, tłumacząca podobieństwa i różnice pomiędzy uczniami różnych klas).

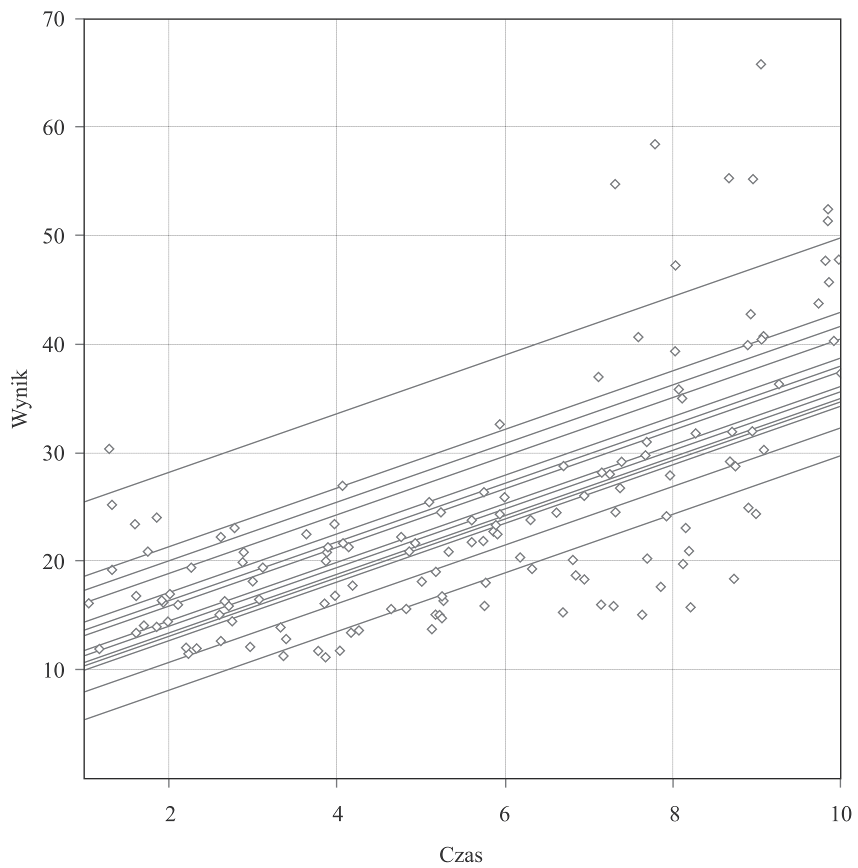
Struktura hierarchiczna z powodzeniem nadaje się też do opisu wyników badań z powtarzanym pomiarem, z tą różnicą, że najniższy poziom stanowić będą powtarzane obserwacje, które traktujemy jako zagnieżdżone w osobach. Osoby traktujemy zaś jak czynnik grupujący drugiego poziomu, mamy więc w takim schemacie liczbę grup równą *n* osób badanych. W modelu hierarchicznym grupy nie muszą być równoliczne, więc każdą z osób może charakteryzować odmienna liczba pomiarów – nie jesteśmy ograniczeni do danych zrównoważonych.

Ponieważ osoby badane stanowią jedynie losową próbkę z interesującej badacza populacji, w modelu regresji wielopoziomowej wprowadzamy czynnik losowy związany z tym międzyosobowym zróżnicowaniem. Dzięki temu, zamiast szukać jednej linii regresji dopasowanej do całego zbioru wyników, dopuszczamy dla każdej z osób linię regresji, która przecina oś *Y* na innej wysokości. W takim modelu możemy testować zarówno efekty zmiennych wewnątrzosobowych (związanych z charakterystyką powtarzanych pomiarów), jak i efekty zmiennych międzyosobowych (charakteryzujących osoby badane) oraz interakcje pomiędzy nimi.

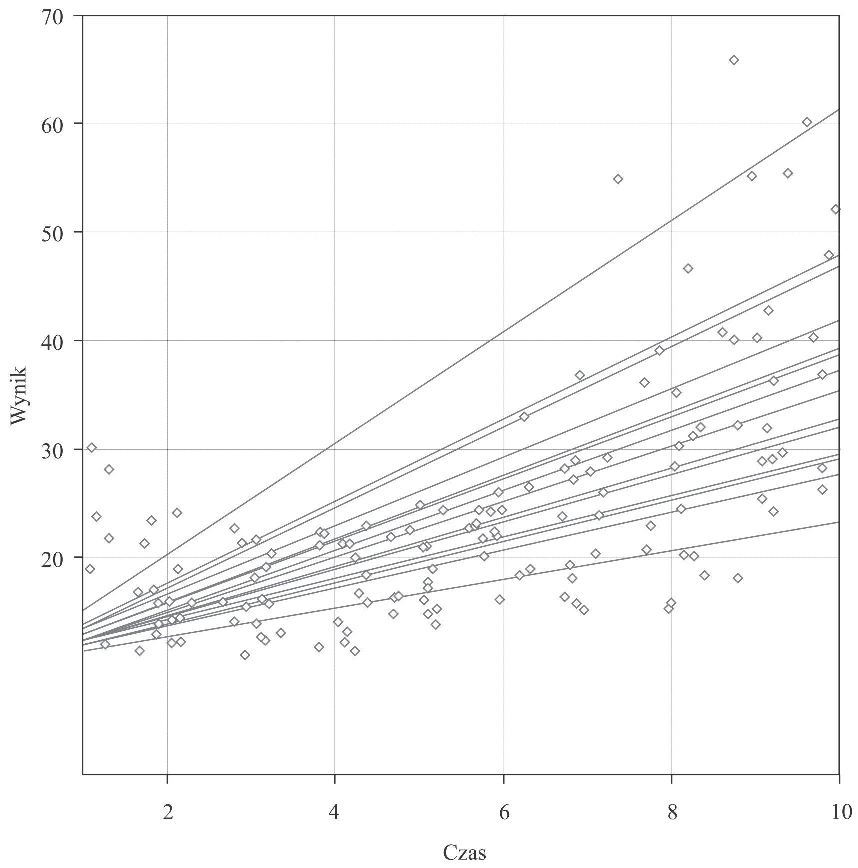
Zmienne wyjaśniające pierwszego poziomu, które opisują sytuacje pomiarowe, mogą mieć charakter efektów stałych, jeśli przyjmujemy, że sytuacje te stanowią zamknięty zbiór interesujących badacza przypadków. W takim wypadku zakładamy, że efekt związany z powtarzanym pomiarem ma u każdej z osób badanych identyczny charakter – linie regresji dla każdej z osób będą wtedy równoległe. Model taki jest analogiczny dla modelu analizy wariancji dla powtarzanych pomiarów (tj. model mieszany z losowym czynnikiem dla osób badanych i stałym czynnikiem dla sytuacji pomiarowych). Nazywa się go czasem modelem efektów losowych dla stałej równania regresji (*random-intercept model*).

Możemy oczekiwać, że efekt związany z powtarzanym pomiarem nie jest stały, ale ma zróżnicowany charakter w populacji osób badanych – najczęściej jest to sensowne przypuszczenie w sytuacji, gdy główną zmienną wewnątrzosobową jest czas pomiaru. W takim wypadku wprowadzamy do modelu dodatkowy składnik wariancji związany ze zmienną pierwszego poziomu, dzięki czemu linie regresji dla poszczególnych osób mogą mieć nie tylko różną wysokość, ale też odmienne nachylenie. Model taki nazywamy modelem z efektem losowym dla nachylenia linii regresji (*random coefficient model; random intercept and slope model*). Graficzne odwzorowanie modeli z czynnikami stałymi i losowymi prezentuje Rysunek 1.

Założenia, jakie musi spełniać hierarchiczny model liniowy, są w pewnym sensie rozszerzeniem założeń klasycznej analizy regresji. Zmienna wynikowa musi być więc zmienną co najmniej przedziałową, reszty pierwszego poziomu muszą mieć rozkład normalny o stałej wariancji we wszystkich grupach, reszty pierwszego poziomu i czynniki losowe wyższych poziomów muszą być nieskorelowane. Ponieważ istotą grupowania hierarchicznego jest podobieństwo w obrębie grup, nie ma założenia niezależności obserwacji na najniższym poziomie hierarchii, jest natomiast założenie niezależności na wyższych poziomach. W przypadku powtarzanych pomiarów oznacza to, iż pomiary dla tej samej osoby mogą być skorelowane, zarazem jednak pomiary dokonywane u poszczególnych osób muszą być od siebie niezależne. Estymacja parametrów modelu hierarchicznego jest dokonywana z reguły za pomocą metody największej wiarygodności (*maximum likelihood*) w jednym z kilku możliwych wariantów. Coraz częściej stosuje się też w tym celu metody typu *bootstrap*. Szczegółowo na temat założeń i metod estymacji można przeczytać w bogatej obecnie anglojęzycznej literaturze – dobrym wprowadzeniem jest m.in. przystępnie napisana książka autorstwa J. Hoxa (1995), w chwili pisania niniejszego artykułu dostępna nieodpłatnie na stronie autora (http://joophox.net/publist/amaboek.pdf). Bardziej zaawansowane pozycje to m.in. Pinheiro i Bates (2000), Raudenbush i Bryk (2002) czy Snijders i Bosker (1999).




*Rysunek 1a.*
Oszacowanie linii regresji dla pomiarów powtarzanych w czasie: hierarchiczny model liniowy z efektem losowym dla stałej równania regresji (oddzielna linia regresji dla każdej z osób).




*Rysunek 1b.*
Oszacowanie linii regresji dla pomiarów powtarzanych w czasie: hierarchiczny model liniowy z efektem losowym dla nachylenia linii regresji (oddzielna linia regresji dla każdej z osób).

## Jak liczyć?

Następujący w ostatnich latach rozwój oprogramowania statystycznego daje badaczowi, który chce w swej praktyce stosować hierarchiczne modele liniowe, wiele możliwych rozwiązań. Istnieją obecnie na rynku wyspecjalizowane aplikacje, służące do prowadzenia analiz wielopoziomowych – dwie najbardziej znane to HLM (http://www.ssicentral.com/hlm/) i MLwiN (http://www.cmm.bristol.ac.uk/MLwiN/). Moduły do analizy liniowych modeli mieszanych znajdują się też w bardziej uniwersalnych pakietach statystycznych, m.in. w pakiecie SAS (http://www.sas.com) czy SPSS (obecnie funkcjonującym pod nazwą PASW Statistics, http://www.spss.pl). Ciekawą i zyskującą coraz większą popularność propozycją jest darmowe środowisko statystyczne *R* (http://www.r-project.org), do którego istnieje obecnie kilka zaawansowanych modułów do prowadzenia analizy wielopoziomowej – podstawowe z nich to nlme (Pinheiro, Bates, DebRoy, Sarkar i R Core team, 2009) i lmer4 (Bates i Maechler, 2009). Wadą *R* jest brak złożonego interfejsu graficznego – sprawne posługiwanie się tym środowiskiem wymaga konstruowania poleceń i skryptów w specjalnie do tego celu stworzonym języku. Zaletą jest jednak ogromna elastyczność w prowadzeniu analiz oraz dostęp do najnowszych, wciąż uaktualnianych procedur statystycznych.

W przypadku powtarzanych pomiarów, zastosowanie hierarchicznego modelu liniowego wymaga wcześniejszego przygotowania danych. Standardowo wyniki

*Tabela 1.*
Przykład wyników badania z powtarzanym pomiarem zapisanych w formacie szerokim

| Osoba | Pomiar 1 | Pomiar 2 | Pomiar 3 | Zmienna opisująca osobę (np. wiek) |
|---|---|---|---|---|
| 1 | 803 | 765 | 922 | 21 |
| 2 | 720 | 727 | 880 | 20 |
| 3 | 838 | 768 | 908 | 24 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| $i$ | $t1_i$ | $t2_i$ | $t3_i$ | $z_i$ |

*Tabela 2.*
Przykład wyników badania z powtarzanym pomiarem zapisanych w formacie długim

| Osoba | Zmienna indeksująca pomiary | Wynik | Zmienna opisująca osobę (np. wiek) |
|---|---|---|---|
| 1 | Pomiar 1 | 803 | 21 |
| 1 | Pomiar 2 | 765 | 21 |
| 1 | Pomiar 3 | 922 | 21 |
| 2 | Pomiar 1 | 720 | 20 |
| 2 | Pomiar 2 | 727 | 20 |
| 2 | Pomiar 3 | 880 | 20 |
| 3 | Pomiar 1 | 838 | 24 |
| 3 | Pomiar 2 | 768 | 24 |
| 3 | Pomiar 3 | 908 | 24 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| $i$ | $t_{1i}$ | $y_{1i}$ | $z_i$ |
| . | . | . | . |
| $i$ | $t_{ji}$ | $y_{ji}$ | $z_i$ |

powtarzanych pomiarów kodowane są w tzw. formacie szerokim, z wynikiem dla każdej z sytuacji pomiarowych wpisanym w oddzielną kolumnę. Przykład tego formatu prezentuje Tabela 1. Ponieważ podstawą dla analizy wielopoziomowej są nie osoby, lecz obserwacje (równo- lub nierównoliczne u każdej z osób), format ten w większości pakietów statystycznych musi być zmieniony na tzw. format długi, w którym wszystkie pomiary tej samej zmiennej zgromadzone są w jednej kolumnie, natomiast charakterystyka pomiaru (u której osoby, w którym momencie itp.) kodowana jest w oddzielnych kolumnach indeksujących. Wartości zmiennych drugiego poziomu, a więc stałych wewnątrzosobowo, powtórzone są w każdym z wierszy charakteryzujących konkretną osobę. Przekształcone na format długi dane z Tabeli 1 prezentuje Tabela 2.

W niniejszym artykule znajdują się wskazówki dotyczące prowadzenia analiz wielopoziomowych za pomocą pakietu SPSS, gdyż jest to pakiet statystyczny dosyć dobrze znany wśród psychologów, a jego niewątpliwą zaletą jest proste w obsłudze środowisko graficzne. Nie da się jednak ukryć, że opcje SPSS dla analiz wielopoziomowych są stosunkowo ograniczone – pakiet ten nie pozwala m.in. na automatyczne porównywanie dopasowania kilku możliwych modeli, dla zmiennych kategorialnych narzuca kodowanie zero-jedynkowe (*dummy coding*), a co najważniejsze nie wspiera analizy uogólnionych liniowych modeli mieszanych (odpowiedniki np. regresji logistycznej czy regresji Poissona dla hierarchicznych modeli liniowych) czy nieliniowych modeli mieszanych. Wszystkie te funkcje dostępne są m.in. w darmowym środowisku *R*, dlatego z myślą o bardziej zaawansowanych Czytelnikach w Załączniku C umieszczone zostały skrypty do *R*, ilustrujące sposób wykonania w tym środowisku przykładowych analiz opisywanych w niniejszym artykule.

Wszystkie zawarte w artykule wykresy zostały wykonane za pomocą pakietu ggplot2 (Wickham, 2009) w wersji 0.8.5 w środowisku statystycznym *R*.

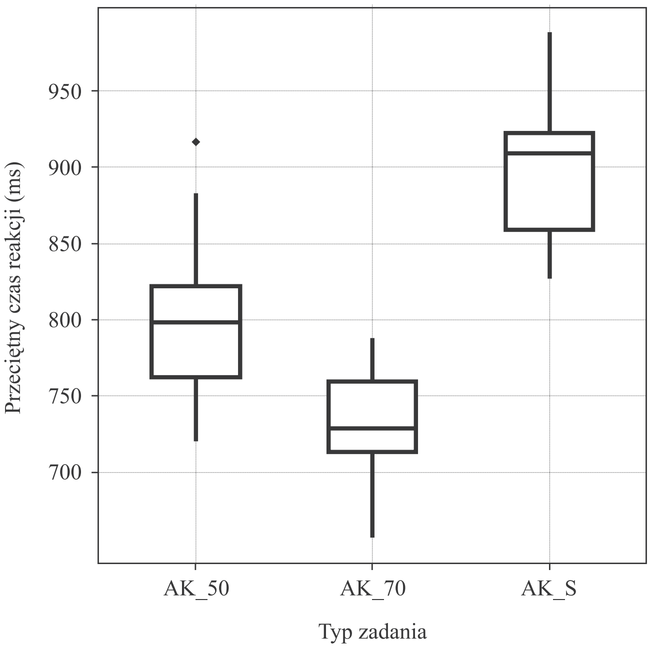
## Przykładowa analiza wyników ze schematu z powtarzanym pomiarem

Pierwszy prezentowany w niniejszym artykule przykład ma na celu zademonstrowanie, że w prostych sytuacjach, gdy spełnione są wszelkie założenia, zarówno ANOVA, jak i model hierarchiczny prowadzą do identycznych wniosków. Analizie poddany został zbiór zrównoważonych danych, w których jedyną zmienną niezależną jest zmienna opisująca trzy rodzaje warunków eksperymentalnych zastosowanych w badaniu. Dane te, opracowane wpierw za pomocą analizy wariancji, następnie zaś w hierarchicznym modelu liniowym, pokazują, że mimo odmiennej metody prowadzenia obliczeń, interpretacja wyników w obu wypadkach przebiega bardzo podobnie. Przestawienie się z klasycznego modelu analizy wariancji na bardziej złożone analizy wielopoziomowe nie powinno więc sprawiać trudności przeciętnie zaawansowanemu w analizie statystycznej badaczowi.

Przykładowy zbiór danych zawiera wyniki z badania aparatem krzyżowym (przyrząd badający koordynację i sprawność reakcji). Dwadzieścia osób zostało poddanych badaniu najpierw w sytuacji tempa wymuszonego (w dwóch wariantach: 50 bodźców na minutę i 70 bodźców na minutę), a następnie w sytuacji tzw. tempa swobodnego (kolejny bodziec pojawia się dopiero po reakcji na poprzedni). Zmienną zależną jest przeciętny (mediana) czas prawidłowej reakcji na sygnał w każdej z trzech sytuacji pomiarowych, wyrażony w milisekundach. Zbiór danych znajduje się w Załączniku A. W tekście artykułu zawarty jest opis, jak przeprowadzić analizę w interfejsie graficznym pakietu SPSS, natomiast w Załączniku C znajduje się skrypt pozwalający przeprowadzić tę samą analizę w środowisku statystycznym *R*.

Rozkłady wyników dla każdego z trzech powtórzonych pomiarów prezentuje Rysunek 2. We wszystkich sytuacjach rozkład wyników nie odbiega istotnie od rozkładu normalnego (w teście Shapiro–Wilka $p$ wynosi odpowiednio dla każdego z warunków: 0,601, 0,405 i 0,357). Aby przeprowadzić analizę wariancji dla powtarzanych pomiarów, potrzebujemy zbioru danych w formacie szerokim (Załącznik A). Po wprowadzeniu danych, w pakiecie SPSS wybieramy polecenie „Analiza → Ogólny model liniowy → Powtarzane pomiary…" i definiujemy czynnik wewnątrzobiektowy *TYP_AK* o trzech poziomach. W kolejnym oknie jako zmienne wewnątrzobiektowe wstawiamy odpowiednio: *AK_50*, *AK_70* i *AK_S*, a w zakładce „Opcje…" zaznaczamy „Oceny wielkości efektu". Możemy tu również zdefiniować pokazywanie średnich dla zmiennej *TYP_AK* i włączyć porównanie efektów głównych z poprawką Bonferroniego dla porównań wielokrotnych. Wykonujemy analizę.

Zbiór danych spełnia podstawowe założenia – zachowany jest warunek sferyczności. W teście Mauchly'ego $W = 0,97$, $p = 0,784$, co oznacza, że nie ma podstaw do odrzucenia hipotezy zerowej o sferyczności w obrębie wyników, analizę wariancji dla powtarzanych pomiarów można więc w tym wypadku uznać za metodę trafną (szersze informacje na temat założenia sferyczności znajdują się w dalszej części artykułu, na razie wystarczająca jest informacja, że zostało ono spełnione). W raporcie SPSS najbardziej interesują nas „Testy efektów wewnątrzobiektowych" – ich podsumowanie prezentuje Tabela 3.




*Rysunek 2.*
Rozkłady wyników dla trzech sytuacji pomiarowych (przykład pierwszy).

Jak widać, test *F* wskazuje na wysoką istotność statystyczną, co oznacza, że wyniki co najmniej jednej z par sytuacji pomiarowych okazały się odmienne. Wynik testu *F* nie wskazuje nam jednak charakteru tych różnic. Można go określić np. za pomocą analizy *post hoc*, w której porównujemy parami wszystkie sytuacje pomiarowe, stosując korektę poziomu istotności ze względu na porównania wielokrotne. Alternatywą dla analizy *post hoc* jest zdefiniowanie kontrastów planowanych. W programie SPSS, przy analizie wariancji z powtarzanym pomiarem możemy wybrać jeden z sześciu wbudowanych w program kontrastów planowanych lub jedną z trzech metod analizy *post hoc*, np. porównania wielokrotne z dosyć konserwatywną poprawką Bonferroniego – wykazują one, że nawet przy uwzględnieniu tej poprawki przeciętny czas reakcji w sytuacji tempa swobodnego ($M = 901{,}7$) był istotnie ($p < 0{,}001$) wyższy zarówno od przeciętnego czasu reakcji w tempie narzuconym 50 bodźców na minutę ($M = 796{,}25$), jak i w tempie narzuconym 70 bodźców na minutę ($M = 736{,}55$). Równocześnie czas reakcji w tempie 50 bodźców na minutę był istotnie ($p < 0{,}001$) wyższy od czasu reakcji w tempie najszybszym (70 bodźców na minutę).

Oprócz istotności statystycznej, ważnym parametrem pokazującym, czy możemy mówić o znaczącym wpływie zmiennej niezależnej na zależną, jest siła efektu. Dla analizy wariancji z powtarzanym pomiarem można obliczyć różne miary siły efektu, jednak najczęściej stosowaną jest $\eta^2_P$ (cząstkowe eta kwadrat). Parametr ten określa proporcję wariancji wyjaśnionej w stosunku do wariancji całkowitej związanej z danym czynnikiem, i w przypadku modelu z jedną zmienną określa po prostu, jaką część wewnątrzosobowej wariancji wyników można przypisać zmiennej niezależnej. Oblicza się go, dzieląc sumę kwadratów dla danej zmiennej przez łączną sumę kwadratów dla tej zmiennej i związanego z nią składnika błędu. W wypadku prezentowanych wyników będzie to więc:

$$\eta^2_p = \frac{279722}{279722 + 27835} = 0{,}909$$

co oznacza dużą siłę efektu – większość (około 90%) wewnątrzosobowego zróżnicowania wyników może być przypisana typowi zadania określającego każdą z trzech sytuacji pomiarowych.

Warto w tym miejscu zaznaczyć, że parametr ten pomija w przypadku powtarzanych pomiarów zmienność międzyosobową, nie jest przez to uniwersalną miarą siły efektu i jego wartość nie może być bezpośrednio porównywana np. z wartością uzyskaną w analogicznym badaniu wykonanym w planie grup niezależnych. Z tego względu w interpretacji i ocenie wielkości $\eta^2_P$ zawsze trzeba uwzględnić charakter i złożoność planu badawczego. W tym jednak wypadku, czyli planu z jednym czynnikiem wewnątrzosobowym, miara ta jest łatwo interpretowalna.

Analiza wariancji pokazuje więc, że najwolniejsze reakcje badani prezentowali w tempie swobodnym, bez presji czasowej. Presja czasowa działała zaś mobilizująco – szybsze zmiany bodźców wymuszały dostosowanie się

*Tabela 3.*
Test efektów wewnątrzosobowych w analizie wariancji dla powtarzanych pomiarów – przykład 1

| | *df* | Suma kwadratów | Średni kwadrat | *F* | $\eta^2_P$ |
|---|---|---|---|---|---|
| TYP_AK | 2 | 279722 | 139861 | 190,94 *** | 0,909 |
| Błąd | 38 | 27835 | 733 | | |

*** $p < 0{,}001$

do tempa i szybsze reakcje. Wszystkie te różnice między sytuacjami są istotne statystycznie, charakterystyka zaś sytuacji pomiarowych tłumaczy niemal całe wewnątrzosobowe zróżnicowanie wyników.

Aby analizować te same wyniki w ramach hierarchicznego modelu liniowego, potrzebujemy zbioru danych w formacie długim. W pakiecie SPSS możemy przekształcić zbiór danych z formatu szerokiego na długi za pomocą polecenia „Dane → Restrukturyzuj dane…”. Po decyzji, czy chcemy zachować oryginalny („szeroki”) zbiór danych w oddzielnym pliku, w pierwszym oknie dialogowym wybieramy pierwszą opcję („Wybrane zmienne przekształcić w obserwacje”), podobnie w drugim, gdyż interesuje nas tylko jeden czynnik wewnątrzosobowy. W trzecim kroku w sekcji „Identyfikacja grup obserwacji” wybieramy „Użyj wybranej zmiennej” i wstawiamy w odpowiednie pole zmienną *OSOBA*. W sekcji „Zmienne do transponowania” zmienną wynikową nazywamy *CZAS_R* i w pole poniżej wstawiamy zmienne *AK_50*, *AK_70* i *AK_S*. W czwartym kroku wybieramy pojedynczą zmienną indeksującą, a w piątym opcję „Nazwy zmiennych”, by nasza zmienna indeksująca zawierała kody oznaczające typ wykonywanego zadania. W polu „Nazwa” zmiennej indeksującej domyślną wartość „Indeks1” zastępujemy nazwą *TYP_AK*. Szósty i siódmy krok zostawiamy w postaci domyślnej, aż do polecenia „Zakończ”.

Dla naszych danych w hierarchicznym modelu liniowym przyjmiemy dwupoziomowy układ, w którym poszczególne pomiary zagnieżdżone są w strukturze wyższego rzędu, charakteryzowanej przez osoby badane. Mamy więc 20 grup wyników, każda z nich zawiera zaś po trzy obserwacje. Przyjmujemy, że osoby badane są losowo dobraną próbką z interesującej nas populacji i że mogą różnić się one poziomem wykonania, ale różnice związane z poszczególnymi sytuacjami są u wszystkich badanych jednakowe – zakładamy więc model z efektem losowym dla stałej równania regresji, a typ zadania traktujemy jako czynnik stały.

Aby rozpocząć estymację modelu, w SPSS wybieramy polecenie „Analiza → Modele mieszane → Liniowe…”. W pierwszym oknie jako „Obiekty” definiujemy zmienną *OSOBA* – jest to w naszym dwupoziomowym modelu jedyny czynnik grupujący obserwacje. Pole „Powtórzone obserwacje” zostawiamy puste – służy ono do tworzenia modelu z narzuconą strukturą macierzy wariancji-kowariancji, co wykracza poza ramy niniejszego tekstu. W kolejnym oknie jako zmienną zależną wstawiamy *CZAS_R*, natomiast w polu „Czynniki” wstawiamy zmienną *TYP_AK*, indeksującą kolejne sytuacje pomiarowe. Ponieważ jest to zmienna kategorialna, wstawienie jej w pole „Czynniki” sprawi, że zostanie ona na potrzeby modelu regresji automatycznie przekodowana na zmienne instrumentalne. SPSS stosuje w tym celu kodowanie zero-jedynkowe (*dummy coding; treatment coding*) z ostatnim z pomiarów traktowanym jako kategorią odniesienia. Warto jednak pamiętać, że nie jest to jedyny sposób kodowania (por. np. Brzeziński, 1996) i w niektórych sytuacjach inne metody mogą dać bardziej interesujące rezultaty. W takim wypadku ograniczenie modułu SPSS można obejść, tworząc samodzielnie w arkuszu danych odpowiednie zmienne instrumentalne i włączając je wszystkie do analizy za pomocą pola „Współzmienne”.

Po powyższych przygotowaniach przystępujemy do budowy modelu. W zakładce „Stałe…” dodajemy do modelu zmienną *TYP_AK*, upewniamy się też, że zaznaczona jest opcja „Uwzględnij wyraz wolny”, by włączyć do modelu stałą równania regresji. W zakładce „Losowe…” wstawiamy zmienną *OSOBA* w pole „Kombinacje”, by określić czynnik grupujący, który posłuży do zdefiniowania efektu losowego. Zaznaczamy też pole „Uwzględnij wyraz wolny”, a pole „Model” pozostawiamy puste, by efekty losowe były konstruowane jedynie dla stałej równania regresji. W zakładce „Statystyki…” zaznaczamy opcję „Oceny parametrów”. Wykonujemy analizę.

Podstawowe parametry modelu, umieszczone w raporcie SPSS pod nagłówkami „Oceny parametrów efektów stałych” oraz „Oceny parametrów kowariancji”, zostały zaprezentowane w Tabeli 4. Jak widać, w części stałej modelu otrzymujemy trzy parametry (dla stałej i dla dwóch zmiennych instrumentalnych kodujących typ zadania), które są odpowiednio średnią wyników w zadaniu z tempem swobodnym (kategoria odniesienia odzwierciedlona przez wartość stałej, czyli 901,7), różnicą między średnią dla pomiaru odniesienia a średnim wynikiem w tempie 50 bodźców na minutę (co oznacza, że średnia w tempie 50/min wynosiła 901,7 – 105,45, czyli 796,25) i różnicą między średnią dla pomiaru odniesienia a średnim wynikiem w tempie 70 bodźców na minutę (tak więc średnia w tempie 70/min wynosiła 901,7 – 165,15, czyli 736,55). Wartości testu $t$ pokazują, że obie różnice są istotne statystycznie. Warto też zwrócić uwagę, że widniejący w raporcie SPSS (pod nagłówkiem „Typ III – Testy efektów stałych”) omnibusowy test $F$ dla czynnika *TYP_AK* daje rezultat identyczny jak w prezentowanej wyżej analizie wariancji, tj. $F_{(2, 38)} = 190{,}94$, $p < 0{,}001$.

Część losowa modelu zawiera wariancję związaną ze stałą regresji, czyli w tym wypadku wariancję międzyosobową, oraz resztę, czyli niewyjaśnioną wariancję na poziomie wewnątrzosobowym. Ponieważ osoby badane traktujemy jako losową próbkę z interesującej nas populacji, można powiedzieć, że wyniki (tj. przeciętne czasy

*Tabela 4.*
Hierarchiczny model liniowy z efektem losowym dla stałej równania regresji – przykład 1

| Parametr | Oszacowanie parametru | Błąd standardowy | *df* | *t* |
|---|---|---|---|---|
| | Efekty stałe: | | | |
| Stała | 901,70 | 9,99 | 31,65 | 90,29*** |
| *Czynniki pierwszego poziomu* | | | | |
| TYP_AK = t. 50/min | –105,45 | 8,56 | 38 | –12,32*** |
| TYP_AK = t. 70/min | –165,15 | 8,56 | 38 | –19,30*** |
| | Parametry losowe: | | | |
| Reszta | 732,51 | 168,05 | | |
| Stała (wariancja) | 1262,39 | 491,99 | | |

*Uwaga*. Jako metodę estymacji zastosowano metodę warunkowej najwyższej wiarygodności (REML).

*** $p < 0,001$

reakcji w milisekundach) osiągane na aparacie krzyżowym w zadaniu z tempem swobodnym tworzą w populacji rozkład normalny o średniej 901,7 (efekt stały dla stałej równania regresji) i odchyleniu standardowym 35,53 (pierwiastek z wariancji dla stałej równania regresji).

Obliczenie miary siły efektu w hierarchicznym modelu liniowym jest sprawą bardziej skomplikowaną niż w przypadku analizy wariancji, jednak w prezentowanym przykładzie można to wykonać bez szczególnych problemów. Model zawiera co prawda dwie składowe wariancji (dla stałej równania regresji i wariancję resztową), jednak w przypadku powtarzanych pomiarów i czynnika wewnątrzosobowego interesuje nas jedynie wariancja resztowa na pierwszym poziomie. Aby ocenić, jaka część wariancji wewnątrzosobowej jest tłumaczona przez typ wykonywanego zadania, należy najpierw oszacować tzw. model zerowy, czyli model zawierający wyłącznie stałą równania regresji (*intercept-only model; baseline model*), bez zmiennych wyjaśniających. W SPSS model taki definiujemy identycznie jak poprzednio, z tą różnicą, że w zakładce „Stałe…" pole „Model" zostawiamy puste. Na początku raportu pojawia się ostrzeżenie („Ostateczna macierz Hessiana nie jest dodatnio skończona (…)"), w tym wypadku możemy je jednak zignorować. Wynika ono z faktu, że międzyosobowy składnik wariancji w analizowanym przykładzie jest bardzo mały i przy standardowych dla SPSS algorytmach szacowania przyjmuje wartość 0. Oszacowanie tego modelu wykazuje, że wartość stałej wynosi 811,5 – jest to średnia ogólna wyników badania. Ponieważ model nie zawiera żadnych predyktorów, składowe wariancji odzwierciedlają wariancję całkowitą wyników. Interesująca nas wewnątrzosobowa wariancja błędu wynosi 6668,32. Znając tę wartość, możemy obliczyć stopień, w jakim wprowadzenie zmiennej *TYP_AK* redukuje tę niewyjaśnioną część zmienności, czyli obliczyć $R^2$ dla zmiennych pierwszego poziomu w modelu hierarchicznym. Tak obliczony parametr jest miarą analogiczną do $R^2$ znanego z klasycznej regresji liniowej, z zastrzeżeniem, że dotyczy on jedynie wariancji wewnątrzosobowej, a nie całego modelu. Obliczyć go można zgodnie ze wzorem:

$$(1) \quad R_1^2 = \frac{Reszta_{\text{model bazowy}} - Reszta_{\text{model oceniany}}}{Reszta_{\text{model bazowy}}}$$

W tym wypadku wartość ta wynosi:

$$\frac{6668,32 - 732,51}{6668,32} = 0,89$$

co daje oszacowanie niemal identyczne jak wartość $\eta^2_p$ w analizie wariancji – typ zadania tłumaczy około 90% wewnątrzosobowej zmienności wyników. Należy pamiętać, że $R^2_1$ jest obliczane łącznie dla wszystkich zmiennych pierwszego poziomu hierarchii i nie wskazuje indywidualnego wkładu poszczególnych czynników w wariancję wyjaśnioną, w odróżnieniu od współczynnika $\eta^2_p$ w analizie wariancji, który jest liczony oddzielnie dla każdej ze zmiennych. Jednak w przypadku modelu z jednym predyktorem obie miary są porównywalne.

Powyższy przykład wykazał, że analiza wykonana w hierarchicznym modelu liniowym daje w przypadku prostego planu badawczego i zrównoważonych danych praktycznie identyczne wyniki jak ANOVA dla powtarzanych pomiarów. Otrzymujemy zarówno te same parametry testu $F$ dla efektów stałych, jak i identyczne wyniki w obrębie różnic między poszczególnymi sytuacjami pomiarowymi – z zastrzeżeniem, że w hierarchicznym modelu liniowym mamy odpowiednik kontrastów planowanych, a nie analizy *post hoc*. Otrzymujemy też niemal identyczne oszacowanie siły efektu dla wprowadzonej zmiennej. Osoby znające metodę analizy wariancji nie powinny mieć więc problemu z interpretowaniem wyników uzyskiwanych za pomocą hierarchicznego modelu liniowego, a korzystanie z nowej metody nie wymaga radykalnej zmiany w stosowanym „paradygmacie badawczym". Pojawia się oczywiście pytanie, gdzie w takim razie leży przewaga liniowych modeli mieszanych, skoro prowadzą do tych samych wniosków, co analiza wariancji. Odpowiedź na to jest prosta – istnieje wiele sytuacji, w których model ANOVA, ze swymi dość konserwatywnymi założeniami, nie powinien (gdyż prowadziłby do obciążonych wyników) lub wręcz nie może być stosowany. W wielu takich przypadkach hierarchiczne modele liniowe wciąż mogą być z powodzeniem stosowane jako trafna metoda analizy danych z powtarzanym pomiarem – opis tych sytuacji wraz z kolejnym przykładem zawierają następne punkty niniejszego artykułu.

## Zyski (i straty) z zastosowania hierarchicznych modeli liniowych do schematu z powtarzanym pomiarem

### Korelacje w obrębie powtarzanych pomiarów

Specyfiką badań z powtarzanym pomiarem zmiennej zależnej jest fakt, że od każdej osoby uzyskujemy nie jeden wynik, lecz zbiór wyników, które są ze sobą skorelowane. Korelacja ta musi więc być kontrolowana w analizie statystycznej prowadzonej na tego typu danych. Znajduje to odzwierciedlenie w założeniu sferyczności (*sphericity*).

Założenie to jest w pewnym sensie rozszerzeniem założenia homogeniczności wariancji, znanego z analizy wariancji dla czynników międzygrupowych. Ponieważ jednak w sytuacji powtarzanych pomiarów testujemy w gruncie rzeczy istotność zmian zachodzących między poszczególnymi sytuacjami pomiarowymi, ważne jest, by wariancje różnic między poszczególnymi parami wyników były homogeniczne. Tak więc np. w badaniu o pięciu sytuacjach pomiarowych ($P_1$, …, $P_5$) wariancje różnic ($P_1 - P_2$, $P_1 - P_5$ itd.) powinny być identyczne – spełnienie tego warunku nazywamy kołowością (*circularity*). Kołowa macierz wariancji-kowariancji, po przekształceniu jej na formę ortonormalną, przyjmuje postać sferyczną (ma równe wariancje i zerowe kowariancje). O wymogu sferyczności możemy mówić, gdy liczba powtarzanych pomiarów jest większa od dwóch, w sytuacji bowiem jedynie dwóch pomiarów mamy tylko jedną różnicę ($P_1 - P_2$). Jeśli założenie sferyczności nie jest spełnione, test $F$ jest pozytywnie obciążony i znacznie zwiększa się ryzyko popełnienia błędu typu I, czyli błędnego odrzucenia hipotezy zerowej.

Założenie sferyczności w sposób najbardziej restrykcyjny jest spełnione w warunkach pełnej addytywności (por. Fidell i Tabachnick, 2003), a więc w sytuacji, w której warunki pomiarowe wpływają na każdą z osób w identycznym stopniu (np. wynik każdej z osób jest w warunku $P_2$ dokładnie o trzy punkty wyższy niż w warunku $P_1$). W takim przypadku wariancja różnic w obrębie dowolnej pary wyników wynosi zero. Addytywność jest jednak mało realna do uzyskania w rzeczywistych badaniach – bardziej realnym i zarazem mniej restrykcyjnym warunkiem, w którym sferyczność jest zachowana, jest sytuacja, w której zarówno wariancje wyników z poszczególnych sytuacji pomiarowych, jak i kowariancje każdej z par poziomów czynnika wewnątrzosobowego (a więc poziomu $P_1$ i $P_2$, $P_2$ i $P_3$, $P_1$ i $P_3$ itd.) są homogeniczne. W takiej sytuacji wariancje różnic między poszczególnymi parami wyników są różne od zera, ale wciąż są sobie równe. Układ wyników w powtarzanych pomiarach, w którym występuje zarówno homogeniczność wariancji, jak i homogeniczność kowariancji, nazywamy symetrią złożoną (*compound symmetry*) – w układzie takim zachowany jest warunek sferyczności i wiele tekstów statystycznych przytacza symetrię złożoną dość nieprecyzyjnie, jako podstawowe założenie analizy wariancji dla powtarzanych pomiarów. Spełnienie warunku symetrii złożonej rzeczywiście wystarcza, nie jest jednak niezbędne, bowiem sferyczność jest warunkiem ogólniejszym i są sytuacje, w których jest ona spełniona mimo braku addytywności czy symetrii złożonej i pozwala na wykonanie nieobciążonych testów $F$.

Jednak nawet sferyczność jest założeniem dość idealistycznym i w niektórych sytuacjach mało realnym. Dotyczy to zwłaszcza badań, w których głównym czynnikiem wewnątrzosobowym jest czas pomiaru – oczekiwanym wręcz efektem jest, że zbliżone czasowo momenty pomiaru będzie cechować wyższa korelacja niż momenty pomiaru bardziej odległe w czasie. Same wariancje także mogą być różne w kolejnych sytuacjach pomiarowych z wynikami wykazującymi np. większe zróżnicowanie pod koniec niż na początku długotrwałych badań (por. Rysunek 1). W takich sytuacjach sferyczność często nie jest zachowana, co znacząco wpływa na trafność testu $F$.

Aby ustalić, czy test $F$ w analizie wariancji z powtarzanym pomiarem można uznać za trafny, założenie sferyczności może być testowane m.in. za pomocą testu Mauchly'ego. Nie jest to rozwiązanie idealne, gdyż test ten wykazuje małą moc w przypadku niewielkich prób, w dużych zaś próbach okazuje się istotny nawet w przypadkach, gdy brak sferyczności ma znikomy wpływ na test *F*. Test Mauchly'ego jest też bardzo wrażliwy na odstępstwa od rozkładu normalnego (por. Davis, 2002). Niemniej jednak w większości pakietów statystycznych jest to standardowa procedura w analizie wariancji dla powtarzanych pomiarów. Jeśli test Mauchly'ego okaże się istotny (co oznacza, że warunku sferyczności nie można uznać za spełniony), dwa podstawowe rozwiązania tej sytuacji to:

1. Korekta stopni swobody w teście *F* za pomocą obliczonego w tym celu współczynnika $\varepsilon$. Jeżeli założenie sferyczności jest spełnione, wartość $\varepsilon$ wynosi 1, jeśli zaś brak sferyczności, wartość $\varepsilon < 1$. Przemnożenie stopni swobody przez $\varepsilon$ pozwala uzyskać wartość *p* zbliżoną do rzeczywistej, tj. nieobciążoną pozytywnie z powodu niespełnienia założeń testu *F*. Najczęściej stosuje się poprawkę Greenhouse–Geissera, uznawaną za dość konserwatywną, tj. lekko zaniżającą wartość $\varepsilon$. Innym rozwiązaniem jest bardziej liberalny wzór Huynh–Feldta (w którym z kolei wartość $\varepsilon$ może być lekko zawyżona i przekraczać 1).

2. Zastosowanie wielozmiennowej analizy wariancji (MANOVA), w której nie ma tak restrykcyjnych założeń dotyczących korelacji między wynikami. Przy dużej liczbie czynników wewnątrzobiektowych MANOVA wymaga jednak stosunkowo dużej liczby osób badanych, by zachować zadowalającą moc testu.

Oba te rozwiązania są automatycznie udostępniane w pakiecie statystycznym SPSS, ich opis można znaleźć np. w podręczniku pod redakcją Bedyńskiej i Brzezickiej (2007). W tym miejscu wystarczy napisać, że obydwa rozwiązania, ze wszystkimi ich wadami i zaletami, w pewnym sensie ukrywają przed badaczem pojawiające się pytanie o to, jaka rzeczywiście jest struktura korelacji charakteryzująca uzyskane wyniki. W niektórych sytuacjach pytanie to może być ważniejsze niż pytanie o istotność różnicy między średnimi z kolejnych pomiarów – jednym z przykładów mogą być badania podłużne nad skutecznością terapii, w których w początkowym okresie u wszystkich pacjentów widoczna jest umiarkowana poprawa, później jednak efekt przestaje być jednorodny. U części osób może nastąpić – w wyniku skutecznej terapii – znaczne zmniejszenie objawów choroby, podczas gdy u innych, niewrażliwych na stosowaną terapię, natężenie choroby stabilizuje się lub nawet zaczyna zaostrzać. W takiej grupie, mimo zwiększania się wariancji wyników, średni poziom natężenia objawów może być zgodny z wcześniej obserwowanym trendem i wykazywać jedynie umiarkowaną poprawę w miarę upływu czasu.

W hierarchicznej regresji liniowej, gdy czynnik wewnątrzosobowy ma charakter zmiennej kategorialnej (jak np. w analizowanym wcześniej badaniu aparatem krzyżowym), model efektu losowego dla stałej równania regresji narzuca podobne ograniczenia, jak analiza wariancji. Składniki wariancji (efekt losowy odzwierciedlający różnice międzyosobowe i reszta, czyli błąd pomiaru) nie są związane z sytuacjami pomiarowymi, przez co macierz wariancji-kowariancji tworzy strukturę symetrii złożonej, która jest specyficzną formą sferyczności. Prostym rozwiązaniem jest włączenie do analizy efektu losowego dla nachylenia linii regresji. Wprowadzamy wtedy składnik losowy związany z sytuacjami pomiarowymi, dzięki czemu dopuszczamy odmienne wariancje w kolejnych pomiarach, a w efekcie – odmienne kowariancje (czyli także korelacje) dla poszczególnych par pomiarów. Założenie sferyczności zostaje w takim modelu zniesione.

Dodatkowo, większość programów statystycznych umożliwia włączenie do liniowego modelu mieszanego z góry założonej przez badacza struktury wariancji-kowariancji, która modeluje związki między powtarzanymi sytuacjami pomiarowymi. Pozwala to nie tylko uwzględnić w modelu, ale też ocenić i szczegółowo opisać strukturę korelacji dla poszczególnych pomiarów. Modelowanie macierzy wariancji-kowariancji (tzw. *covariance pattern models*) jest bardziej zaawansowaną techniką i wykracza poza ramy niniejszego artykułu.

**Brakujące obserwacje**

W modelu badawczym z powtarzanymi pomiarami często spotykanym problemem są braki danych. Jest to szczególnie uciążliwe w długo trwających badaniach, w których głównym czynnikiem jest czas dokonywania pomiaru. „Wypadanie" osób z badania prowadzi do niezrównoważonych wyników, co ma konsekwencje dla możliwości prowadzenia trafnych analiz statystycznych i przeprowadzania nieobciążonych testów istotności.

W literaturze (por. Little i Rubin, 1987) przyjęło się rozróżnienie na kompletnie losowe braki danych (*missing completly at random,* MCAR), losowe braki danych (*missing at random,* MAR) i braki nielosowe (*missing not at random,* MNAR). Pierwszy przypadek dotyczy sytuacji, w której „wypadnięcie" osoby jest niezwiązane z wartościami mierzonych zmiennych (np. awaria aparatury pomiarowej). Druga sytuacja następuje wówczas, gdy brak danych nie jest skorelowany z wartością brakującej zmiennej, może być jednak skorelowany z wcześniejszymi pomiarami (np. brak danych o stanie zdrowia pacjenta z powodu braku kontaktu z badanym). W trzeciej sytuacji występuje korelacja między brakiem danych a wartością brakującej zmiennej (np. nieudzielenie w ankiecie odpowiedzi na zbyt osobiste pytanie). Analiza wariancji dla powtarzanych pomiarów zachowuje trafność jedynie w przypadku, gdy braki danych są kompletnie losowe (MCAR). Jednak nawet wtedy pozostaje problem w postaci niezrównoważenia danych, gdyż poszczególne osoby różnią się między sobą liczbą dokonanych pomiarów. Standardowym, choć mającym wady rozwiązaniem jest w takim wypadku prowadzenie analizy jedynie na osobach z kompletnymi wynikami (*complete cases analysis*). Przyczynia się ono jednak do zmniejszenia mocy przeprowadzanych testów i jest szczególnie uciążliwe w przypadku skomplikowanych badań prowadzonych na małych próbach. Inne, bardziej złożone rozwiązania to np. estymowanie wartości brakujących pomiarów lub przeprowadzenie analizy wariancji dla danych niezrównoważonych (*ANOVA for unbalanced design*), mają one jednak charakter metod bardziej zaawansowanych i niepolecanych niedoświadczonym użytkownikom.

W analizie wielopoziomowej, w której szacowanie parametrów modelu opiera się na metodach wiarygodności, warunek losowości jest bardziej liberalny. Przy założeniu, że braki danych nie zależą od wartości brakującej zmiennej (MCAR lub MAR) analiza taka jest traktowana jako trafna. Różna liczba obserwacji u poszczególnych osób nie stanowi problemu, nie ma bowiem w hierarchicznym modelu liniowym wymogu równej liczebności grup. Można więc wykorzystać w analizie wszystkie zebrane wyniki – ma to pozytywny wpływ na moc wykonywanych testów.

Oczywiście, opisane wyżej właściwości nie czynią z hierarchicznego modelu liniowego uniwersalnej metody radzenia sobie z brakami danych w zebranych wynikach. W przypadku dużej ich liczby lub ich nielosowej struktury (gdy na przykład większość braków przypada w powtarzanych pomiarach na ostatnią obserwację) dane przestają być bowiem reprezentatywne dla analizowanej populacji, co znacząco wpływa na możliwość interpretowania i generalizowania uzyskanych wyników – w takim wypadku wybór metody statystycznej (analizy wielopoziomowej, analizy wariancji itp.) jest wtórny wobec znacznie ogólniejszego problemu. Jednak w przypadku sporadycznych braków danych (np. losowo rozłożone, pojedyncze brakujące obserwacje u niewielkiej liczby osób) hierarchiczny model liniowy pozwala wykorzystać w analizie wszystkie zebrane wyniki i nie wymaga wyłączania np. osoby, u której w jednym z kilkunastu analizowanych punktów czasowych nie udało się dokonać pomiaru.

**Rozkład sytuacji pomiarowych w czasie**

Inna wrażliwość analizy wariancji i hierarchicznego modelu liniowego na niezrównoważone dane ma duże znaczenie już przy planowaniu samego schematu badawczego, z którego wyniki będą poddawane analizie tymi metodami. W standardowej analizie wariancji, zakładającej dane zrównoważone, wszystkie osoby muszą być badane w tych samych, stałych wartościach zmiennej niezależnej. Jeśli głównym czynnikiem niezależnym jest czas (co w przypadku powtarzanych pomiarów bywa dosyć częste), oznacza to, że liczba pomiarów i odstępy między poszczególnymi pomiarami u każdej z osób powinny być identyczne.

W hierarchicznym modelu liniowym, który nie jest ograniczony do danych zrównoważonych, badacz nie jest zmuszony do planowania jedynie stałych sytuacji pomiarowych – możliwe jest włączenie do modelu zarówno pomiarów w jednakowych (*fixed occasions*), jak i w różnych momentach (*varying occasions*) u poszczególnych osób. Model regresji wielopoziomowej pozwala, a wręcz sugeruje wprowadzanie zmiennych – a więc i czasu pomiaru – jako zmiennych ilościowych, gdy tylko jest to możliwe. W takiej sytuacji czas pomiaru, traktowany jako zmienna ciągła, może dla każdej z osób przyjmować odmienne wartości, a odległość między poszczególnymi momentami pomiaru jest precyzyjnie ujęta w analizie danych w celu określenia najlepiej dopasowanej linii trendu.

Kolejną ważną cechą hierarchicznych modeli liniowych jest fakt, że czas może być wprowadzony do analizy nie jako efekt stały, ale jako efekt losowy – mamy wtedy do czynienia z modelem efektu losowego dla nachylenia linii regresji. Oznacza to, że dopuszczamy dla każdej z osób badanych inną wielkość zmian w obrębie zmiennej zależnej w czasie przebiegu badania. Dla każdej osoby możemy więc wyznaczyć linię trendu o innych parametrach – różniącą się nie tylko punktem początkowym (efekt losowy dla stałej równania regresji), ale też szybkością (przy losowym trendzie liniowym) czy przyspieszeniem (przy losowym trendzie kwadratowym) spadku lub wzrostu wartości mierzonej zmiennej. Wszystkie te właściwości dają analizie wielopoziomowej znaczącą przewagę nad klasyczną analizą wariancji dla powtarzanych pomiarów.

**Możliwości rozbudowy analizowanego modelu**

W analizie wariancji z powtarzanym pomiarem istnieją oczywiście możliwości rozbudowy podstawowego schematu. Prosty model z jednym czynnikiem wewnątrzosobowym, prezentowany w pierwszym przykładzie, można rozszerzyć, włączając dodatkowe czynniki wewnątrzosobowe charakteryzujące warunki badawcze w kolejnych pomiarach (np. do pierwszego przykładu mogliśmy wprowadzić zmienną w postaci liczby prezentowanych bodźców, równej 50 i 100 – otrzymalibyśmy wtedy sześć sytuacji pomiarowych w planie $3 \times 2$, czyli trzy typy zadania na aparacie krzyżowym, każde wykonane dwa razy: raz w krótkiej, a raz w długiej wersji). Można też wprowadzać czynniki międzyosobowe w postaci zmiennych kategorialnych, dzielących badanych na kilka grup (np. osoby, które musiały wykonywać zadanie lewą ręką, i osoby wykonujące je ręką prawą). Oprócz efektów głównych, testujemy wtedy istotność interakcji między czynnikami wewnątrz- i międzyosobowymi.

W porównaniu z analizą wariancji, liniowe modele mieszane dają znacznie większe możliwości rozbudowy. Dodatkowe zmienne objaśniające mogą być wprowadzane na każdym z poziomów modelu. Podobnie jak w analizie wariancji, mogą to być więc dodatkowe zmienne kategorialne wewnątrz- i międzyosobowe oraz ich interakcje, mogą być to także kowarianty (tj. zmienne ciągłe, pozwalające przewidywać zmienną zależną) charakteryzujące osoby badane – będą to zmienne drugiego poziomu. Mamy też możliwość włączenia do analizy zmiennych ciągłych na poziomie pierwszym, czyli tzw. kowariantów zmiennych czasowo (*time varying explanatory variables*), czego nie oferuje klasyczny model ANOVA. Zmiennym czasowo kowariantem może być np. czas poświęcony na wykonanie zadania, gdy główną zmienną zależną jest wynik, czyli poprawność, a niezależną – typ zadania. Może to być też np. lęk rozumiany jako stan mierzony oddzielnie przed każdą z sytuacji pomiarowych.

Zmienny czasowo kowariant daje także badaczowi możliwość uwzględnienia w analizie statystycznej równocześnie efektu czasu i efektu związanego ze specyfiką sytuacji pomiarowej (np. trudnością zadania czy typem prezentowanych bodźców) – ma to oczywiście sens jedynie w sytuacji, gdy czas i typ sytuacji pomiarowej nie są skorelowane, a więc np. w eksperymencie, w którym każdemu z badanych prezentowano bodźce eksperymentalne w odmiennej kolejności.

Hierarchiczny model liniowy dla powtarzanych pomiarów, podobnie jak każdy inny model wielopoziomowy, nie jest ograniczony do dwóch poziomów. Daje to dodatkowe możliwości rozbudowy analizowanego modelu poprzez rozszerzenie jego struktury hierarchicznej. Najniższym poziomem standardowo jest poziom obserwacji, zaś poziom drugi to poziom charakteryzujący osoby badane. Można sobie jednak łatwo wyobrazić, że osoby badane są jednostkami pochodzącymi z nadrzędnej hierarchicznej struktury – są to np. losowo wybrani uczniowie z różnych szkół. Czynnik grupujący osoby badane będzie wtedy tworzył poziom trzeci, który może oczywiście zawierać zmienne objaśniające właściwe dla tego poziomu (np. wielkość szkoły, lokalizacja w małym/dużym mieście itp.) i wprowadzać dodatkowe interakcje między tymi zmiennymi a zmiennymi poziomów niższych.

**Niedogodności związane z interpretowaniem parametrów w analizie wielopoziomowej**

Duża elastyczność liniowych modeli mieszanych jest ich zdecydowaną zaletą, sprawne wykorzystanie tej elastyczności może jednak początkowo sprawiać trudność badaczom przywiązanym do schematu analizy wariancji. W analizach wielopoziomowych pojawiają się bowiem wyzwania znane z klasycznej analizy regresji, dodatkowo komplikowane przez hierarchiczną strukturę danych. Jest to szczególnie widoczne przy próbach interpretacji wartości wyrażenia stałego w równaniu regresji.

W analizie wariancji zmienne wewnątrzosobowe mają charakter zmiennych kategorialnych, nawet gdy pierwotnie były wyrażone na skali wyższej. W liniowym modelu mieszanym najlepszym rozwiązaniem jest wprowadzanie zmiennych ilościowych, jeśli jednak jesteśmy zmuszeni uwzględnić czynniki o charakterze nominalnym, muszą one zostać wprowadzone do równania regresji za pomocą odpowiednio zakodowanych zmiennych instrumentalnych. Zastosowanie równocześnie kilku zmiennych o wielu kategoriach (a więc takich, które wymagają więcej niż jednej zmiennej instrumentalnej) i modelowanie interakcji między nimi staje się w takim wypadku szczególnie kłopotliwe interpretacyjnie.

Kolejną kwestią dobrze znaną w analizie regresji, a niemającą szczególnego znaczenia w analizie wariancji jest centrowanie zmiennych. W przypadku zmiennych niecentrowanych, wyrażenie stałe równania regresji przyjmuje wartość, jaką zgodnie z modelem uzyska zmienna przewidywana, gdy wszystkie pozostałe zmienne będą równe zeru. W niektórych sytuacjach jest to wartość bezsensowna, np. gdy jednym z predyktorów był Iloraz Inteligencji (ze średnią 100 i odchyleniem standardowym równym 15). Centrowanie jest zabiegiem przesuwającym punkt zerowy; np. centrowanie wokół średniej wartości I.I. w badanej próbie sprawi, że wyrażenie stałe będzie określać poziom zmiennej przewidywanej u badanego o przeciętnym poziomie inteligencji w danej grupie, a np. centrowanie wokół wartości 100 sprawi, że wartość stałej regresji odzwierciedli oczekiwany wynik przeciętnej osoby w populacji ogólnej. Podobnie jak w przypadku kodowania zmiennych kategorialnych, centrowanie wymaga od badacza szczególnej uwagi przy interpretowaniu parametrów równania regresji w wypadku występowania interakcji między zmiennymi z różnych poziomów hierarchii.

**Miary siły efektu w liniowych modelach mieszanych**

Istotnym problemem jest kwestia oceny siły efektu dla testowanych zmiennych, poruszana już przy okazji analizy pierwszego przykładu. W analizie wariancji funkcjonują różne wskaźniki (*f* Cohena, $\eta^2$, $\eta^2_{\mathrm{P}}$, czy zaproponowany w ostatnich latach jako najbardziej uniwersalny $\eta^2_G$, czyli zgeneralizowane eta-kwadrat – por. Olejnik i Algina, 2003), które pozwalają ocenić, czy wpływ testowanych czynników na zmienną przewidywaną jest nie tylko nieprzypadkowy (czyli istotny statystycznie), lecz także mający realne znaczenie (czyli charakteryzujący się dużą siłą efektu, co z reguły oznacza, że czynnik ten odpowiada za duży procent wariancji zmiennej wyjaśnianej). Wiele czasopism wymaga obecnie, by – cytując wyniki analizy statystycznej – podawać nie tylko istotność, ale też miarę siły efektu właściwą dla danego testu statystycznego.

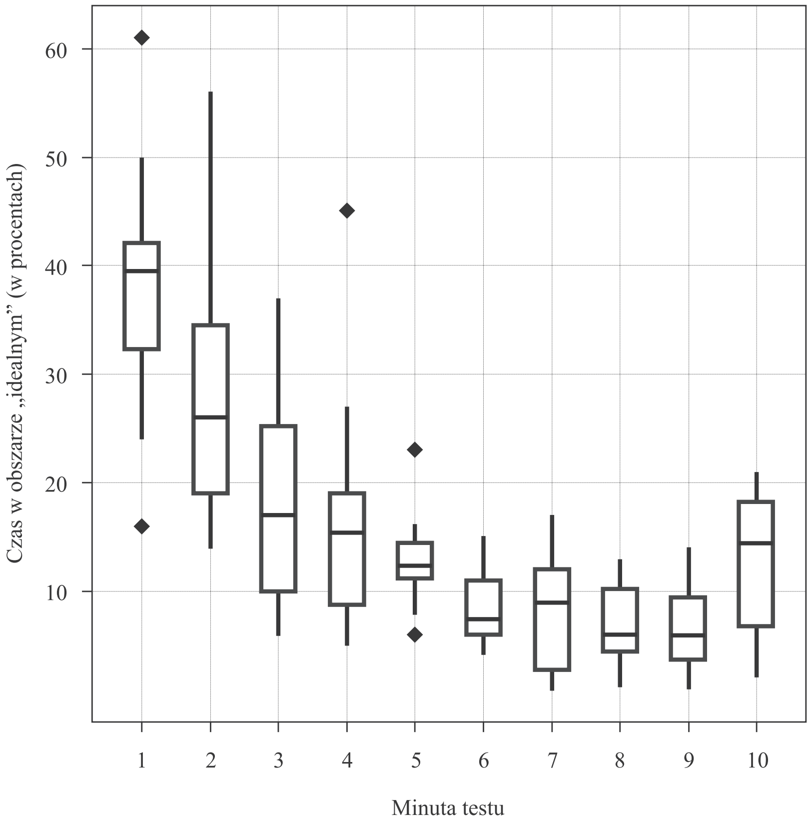
W hierarchicznym modelu liniowym określanie siły efektu nie jest, niestety, proste. Dzieje się tak ze względu na hierarchiczną strukturę danych i rozbicie wariancji na kilka oddzielnych elementów składowych, związanych z poszczególnymi poziomami. W praktyce więc proporcję wyjaśnianej wariancji należy określać oddzielnie dla każdego z tych elementów. Szczegółowo opisują to Snijders i Bosker (1999), przyjmując za miarę wyjaśnianej przez model wariancji przybliżone $R^2$, będący współczynnikiem analogicznym do $R^2$ w klasycznym modelu regresji. Obliczanie takiego współczynnika dla pierwszego poziomu (a więc określenie efektu wewnątrzosobowego w powtarzanych pomiarach) było już opisane w analizie pierwszego przykładu w niniejszym artykule. Przybliżone $R^2$ jest tak naprawdę współczynnikiem określającym redukcję wariancji błędu w testowanym modelu względem modelu bazowego. W analizowanym wcześniej przykładzie za model bazowy posłużył tzw. model zerowy, w którym nie było żadnych zmiennych wyjaśniających, jednak nie w każdej sytuacji będzie to właściwe. Czasem za model bazowy służyć może model, który sam wyjaśnia jakiś procent wariancji (np. dla badań podłużnych będzie to model z wprowadzonym do równania trendem czasowym). W takim przypadku obliczony współczynnik ma w zasadzie charakter $\Delta R^2$, co zmienia nieco jego interpretację. Większy problem stanowi obliczanie przybliżonego $R^2$ dla składowych wariancji związanych z poziomami wyższymi, a więc np. wielkości wyjaśnionej wariancji międzyosobowej. Ze względu na specyfikę dzielenia wariancji na części składowe w modelu hierarchicznym, po wprowadzeniu niektórych predyktorów $\Delta R^2$ dla poziomu drugiego może nawet przyjmować wartość ujemną. Ponadto wraz ze wzrostem skomplikowania modelu i wprowadzaniu wielu zmiennych na różnych poziomach oraz interakcji między nimi, wybór składnika wariancji (bądź ich sumy) stanowiącego podstawę do obliczania $\Delta R^2$ oraz właściwa interpretacja $\Delta R^2$ stają się jeszcze bardziej złożone. Zasadniczo należy przyjąć, że niestety nie ma obecnie prostego i uniwersalnego sposobu oceny siły efektu dla wszystkich czynników w liniowym modelu mieszanym.

### Analiza wyników ze schematu z powtarzanym pomiarem – przykład drugi

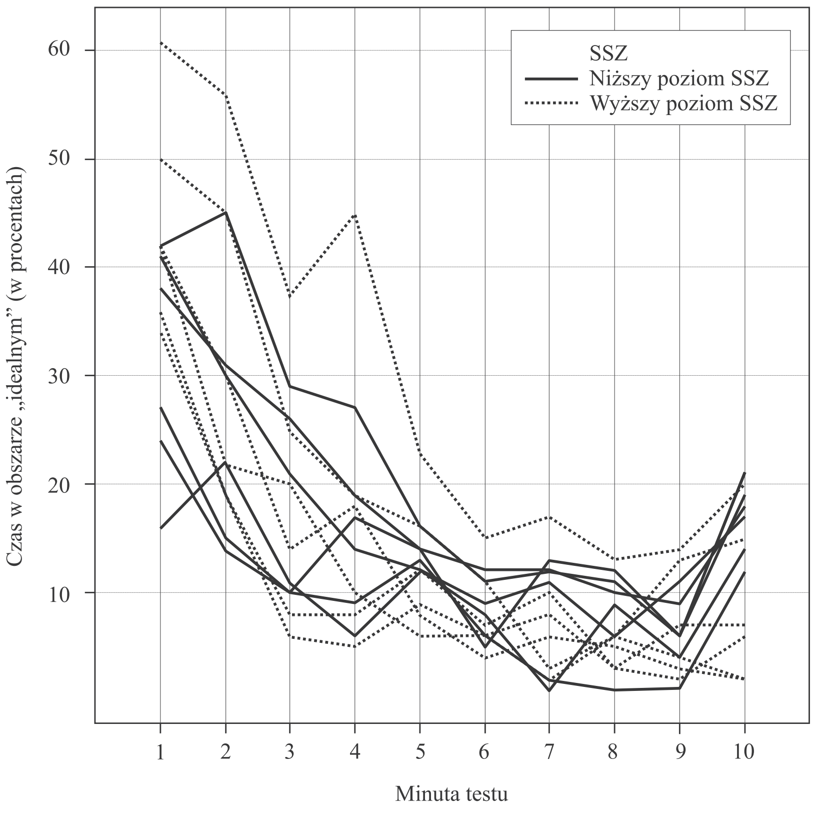
Drugi przykład prezentuje, krok po kroku, analizę danych z czasem pomiaru jako główną zmienną niezależną. Tak jak poprzednio, dane zostały opracowane zarówno w modelu analizy wariancji, jak i za pomocą hierarchicznego modelu liniowego. Mają one charakter zrównoważony i nie ma w nich brakujących obserwacji, co dopuszcza zastosowanie analizy wariancji, wyniki nie spełniają jednak założeń niezbędnych dla stosowania tej metody. W takim wypadku hierarchiczny model liniowy pozwala na trafniejsze wnioskowanie statystyczne – poniższa analiza prezentuje kolejne etapy budowy właściwego modelu dla prezentowanych wyników.

Dwunastu badanych wykonywało test *SMK* z Wiedeńskiego Systemu Testów (http://www.schuhfried.at). *SMK* służy do badania koordynacji i polega na utrzymywaniu, za pomocą odpowiedniej konsoli, widocznego na ekranie ruchomego obiektu w stabilnej pozycji. Podstawowym wskaźnikiem jest czas, w jakim udaje się badanemu realizować to zadanie, czyli utrzymywać uciekający obiekt w obszarze idealnym. Zgromadzone dane dotyczą pierwszych 10 minut testu – dla każdej z minut wyliczono oddzielny wskaźnik (czas utrzymywania obiektu w obszarze idealnym w danej minucie testu, wyrażony w procentach), co pozwala ocenić, jak zmieniał się poziom wykonania podczas tego okresu. Czas pomiaru ma więc w tym przykładzie charakter czynnika wewnątrzosobowego. Każdego z badanych charakteryzuje też wynik w skali Styl Skoncentrowany na Zadaniu (SSZ) z kwestionariusza CISS (Strelau, Jaworowska, Wrześniewski i Szczepaniak, 2005), w tym wypadku mogący służyć jako predyktor determinacji do działania, cechującej badanego w sytuacjach stresujących (a do takich zalicza się test *SMK*). Jest to czynnik międzyosobowy. Wynik ten ma postać zmiennej ilościowej, jednak dla uproszczenia został on zdychotomizowany (na podstawie mediany) i przekształcony w zmienną dzielącą badanych na dwie równoliczne grupy – z wyższym i niższym poziomem SSZ. Zbiór danych znajduje się w Załączniku B. W artykule znajdują się wskazówki, jak prowadzić opisaną analizę w interfejsie graficznym SPSS, natomiast w Załączniku C znajduje się skrypt pozwalający wykonać identyczną analizę w środowisku statystycznym *R*.

Rysunek 3 prezentuje rozkłady wyników w kolejnych minutach testu. Widać na nim wyraźnie, że zmianie ulega nie tylko poziom wykonania, który szybko obniża się w początkowych minutach. Maleje także zróżnicowanie




*Rysunek 3.*
Rozkład wyników w kolejnych sytuacjach pomiarowych (przykład drugi).




*Rysunek 4.*
Indywidualne przebiegi czasowe dla poszczególnych osób (przykład drugi).

wyników, co sugeruje, że mimo początkowych różnic indywidualnych, w dalszej części testu poziom wykonania u wszystkich badanych był zbliżony. Potwierdza to Rysunek 4, prezentujący indywidualne przebiegi czasowe dla wszystkich 12 badanych, z dodatkowym zaznaczeniem poziomu SSZ.

Analiza wariancji w SPSS wymaga tabeli danych w formacie szerokim. Po wprowadzeniu danych wybieramy polecenie „Analiza → Ogólny model liniowy → Powtarzane pomiary…" i definiujemy czynnik wewnątrzobiektowy *POMIAR* o dziesięciu poziomach. W następnym oknie jako zmienne wewnątrzobiektowe wstawiamy zmienne charakteryzujące kolejne pomiary, a w polu „Czynniki międzyobiektowe" wstawiamy zmienną *SSZ*. W zakładce „Opcje…" zaznaczamy „Oceny wielkości efektu". Wykonujemy analizę.

Próba wykonania analizy wariancji z powtarzanym pomiarem dla wewnątrzosobowego efektu czasu (10 sytuacji pomiarowych) i międzyosobowej zmiennej charakteryzującej poziom SSZ nie wygląda zadowalająco. Testowanie założeń wykazuje, że dane te nie spełniają warunku sferyczności, co widać w wyniku testu Mauchly'ego ($W < 0{,}000$, $p < 0{,}000$). Jak było to opisane wcześniej, w takich sytuacjach rozwiązaniem może być korekta stopni swobody testu $F$ zgodnie z wartością $\varepsilon$ (w tym wypadku wartość ta jest bardzo niska – współczynnik Greenhouse–Geissera wynosi 0,231) lub zastosowanie modelu wielozmiennowego (MANOVA). Oba podejścia prowadzą do nieco odmiennych rezultatów, co sugeruje, że struktura danych nie daje się dobrze opisać przynajmniej przez jedną z tych metod. Analiza wielozmiennowa, w której założenie sferyczności nie jest konieczne, wykazuje istotny efekt główny czasu pomiaru, $\Lambda$ *Wilksa* $= 0{,}005$, $F_{(9;2)} = 46{,}008$, $p = 0{,}021$. Interakcja czasu pomiaru z wartością SSZ jest na poziomie tendencji statystycznej, $\Lambda$ *Wilksa* $= 0{,}017$, $F_{(9;6)} = 12{,}918$, $p = 0{,}074$. Wyniki jednozmiennowej analizy wariancji prezentuje Tabela 5. Efekt wewnątrzosobowy związany z sytuacją pomiarową jest istotny (około 82% zmienności wewnątrzosobowej wiąże się z czasem pomiaru). Interakcja czasu pomiaru i poziomu SSZ jest istotna przy $df = 9$, jednak niespełnienie wymogu sferyczności nie pozwala traktować tego wyniku z pełnym zaufaniem. Korekta stopni swobody na podstawie współczynnika $\varepsilon$ lokuje tę interakcję na granicy istotności – przy konserwatywnej poprawce Greenhouse–Geissera osiąga ona jedynie poziom tendencji statystycznej ($p = 0{,}064$), zaś przy bardziej liberalnym współczynniku $\varepsilon$ Huynh–Feldta osiąga kryterium istotności ($p = 0{,}041$). Wykonana dodatkowo analiza trendów (kontrasty wielomianowe) wskazuje, że istotne są jedynie trend liniowy ($F_{(1;10)} = 64{,}587$, $p < 0{,}001$, $\eta^2_p = 0{,}87$) i kwadratowy ($F_{(1;10)} = 163{,}574$, $p < 0{,}001$, $\eta^2_p = 0{,}94$) dla efektu głównego czasu pomiaru. Istotny trend kwadratowy wskazuje, że związek czasu pomiaru z poziomem wykonania nie jest prostoliniowy, ale ma kształt paraboli lub wycinka paraboli. Charakter trendu czasowego i interakcji z poziomem SSZ zmuszeni jesteśmy określić na podstawie wykresów – Rysunek 5 podpowiada, że poziom wykonania obniża się gwałtownie w pierwszych minutach testu, osiągając minimum około ósmej minuty. W ostatniej minucie następuje ponowny wzrost poziomu wykonania. U osób z wyższą wartością SSZ można zaobserwować początkowo wyższy poziom wykonania i gwałtowniejszy spadek na początku testu, mniej jest też u nich widoczny ponowny wzrost wyników.

Szczegółowego opisu różnic wewnątrz- i międzygrupowych dla poszczególnych pomiarów mogą dostarczyć nam testy *post hoc*, jednak w przypadku przedstawionych danych zastosowanie hierarchicznego modelu liniowego jest zdecydowanie lepszym wyborem. Po pierwsze, pozwoli on wprowadzić czas jako zmienną ilościową, co w tym wypadku ma uzasadnienie, po drugie, pozwoli lepiej opisać różnice międzyosobowe nie tylko w poziomie wykonania, ale też w szybkości jego zmiany.

Aby przeprowadzić analizę w pakiecie SPSS, musimy przekształcić dane z formatu szerokiego za pomocą polecenia „Dane → Restrukturyzuj dane…". Wykonujemy to w sposób analogiczny jak w pierwszym przykładzie. W kroku trzecim „Zmienną wynikową" nazywamy *WYNIK*, zaś w pole poniżej wstawiamy zmienne opisujące kolejnych dziesięć sytuacji pomiarowych. Ponieważ tym razem zbiór danych zawiera też czynnik międzyosobowy, w pole „Zmienne niezmienione" wstawiamy zmienną *SSZ*. W kroku piątym zostawiamy domyślne indeksowanie w postaci liczbowej (1, 2, 3 …), zaś zmienną indeksującą nazywamy *POMIAR*. Po zakończeniu restrukturyzacji otrzymujemy tabelę danych w formacie długim.

Przed przystąpieniem do estymacji modelu warto poczynić pewne przygotowania. Po pierwsze, dziesięć sytuacji pomiarowych najlepiej zakodować wartościami od 0 do 9. Najłatwiej wykonać to poleceniem „Przekształcenia → Oblicz wartości…" i zmienną wynikową *POMIAR* zdefiniować jako „*POMIAR* – 1". Dzięki temu wartość stałej równania regresji będzie przeciętnym poziomem wykonania w pierwszej minucie testu, pod warunkiem, że także pozostałe zmienne w modelu przyjmują wartość zerową. Ponieważ do modelu chcemy włączyć zmienną kategorialną *SSZ*, zostanie ona za pomocą kodowania zero-jedynkowego zamieniona na zmienną instrumentalną. SPSS przypisuje automatycznie wartość zero ostatniej kategorii, dzięki czemu stałą równania regresji można

*Tabela 5.*
Test efektów wewnątrzosobowych w analizie wariancji dla powtarzanych pomiarów – przykład 2

| | *df* | Suma kwadratów | Średni kwadrat | *F* | $\eta^2_P$ |
|---|---|---|---|---|---|
| POMIAR | 9 | 11446,375 | 1271,819 | 45,127*** | 0,819 |
| *ε (Greenhouse–Geisser)* | *2,08* | *11446.375* | *5511,114* | *45,127**** | *0,819* |
| *ε (Huynh–Feldt)* | *2,89* | *11446,375* | *3956,197* | *45,127**** | *0,819* |
| POMIAR · SSZ | 9 | 801,842 | 89,094 | 3,161** | 0,24 |
| *ε (Greenhouse–Geisser)* | *2,08* | *801,842* | *386,065* | *3,161* | *0,24* |
| *ε (Huynh–Feldt)* | *2,89* | *801,842* | *277,14* | *3,161** | *0,24* |
| Błąd | 90 | 2536,483 | 28,183 | | |
| *ε (Greenhouse–Geisser)* | *20,77* | *2536,483* | *122,125* | | |
| *ε (Huynh–Feldt)* | *28,93* | *2536,483* | *87,668* | | |

*Uwaga. W Mauchly'ego* < 0,001, *p* < 0,001; *ε Greenhouse-Geissera* = 0,231; *ε Huynh-Feldta* = 0,321.

* *p* < 0,05; ** *p* < 0,01; *** *p* < 0,001

będzie interpretować jako przeciętny poziom wykonania testu *SMK* u osoby z wyższym poziomem SSZ.

Z wykresów widać, że zmiana wykonania testu *SMK* w czasie nie ma charakteru prostoliniowego – oprócz zmiennej *POMIAR* potrzebujemy więc także zmiennej *POMIAR2*, zawierającej wartości zmiennej *POMIAR* podniesione do kwadratu, aby wprowadzić ją do modelu (samodzielnie zmienną taką można stworzyć poleceniem „Przekształcenia → Oblicz wartości…"; jako zmienną wynikową wpisujemy *POMIAR2*, którą definiujemy jako *POMIAR · POMIAR*).

Po tych przygotowaniach, pierwszym krokiem, jaki powinniśmy wykonać, jest oszacowanie modelu zerowego, który będzie nam służył za model bazowy dla dalszych porównań. W SPSS wybieramy więc polecenie „Analiza → Modele mieszane → Liniowe…". W pierwszym oknie jako „Obiekty" wstawiamy zmienną *OSOBA*. W kolejnym oknie jako zmienną zależną wstawiamy *WYNIK*. Już w tym momencie możemy w polu „Czynniki" wstawić zmienną kategorialną *SSZ*, opisującą różnice indywidualne, a w polu „Współzmienne" wstawić zmienne ilościowe *POMIAR* i *POMIAR2*, które posłużą do modelowania czasu pomiaru. W zakładce „Statystyki…" zaznaczamy opcję „Oceny parametrów", zaś w zakładce „Estymacja…" zmieniamy metodę szacowania na „Największą wiarygodność (ML)" – pozwoli to nam w sposób bezpośredni porównywać różniące się od siebie modele (więcej na ten temat w dalszej części przykładu). Po tych przygotowaniach, kolejne modele będziemy definiować, modyfikując jedynie zakładki „Stałe…" i „Losowe…" – pozostałe opcje zostawimy bez zmian.

Model zerowy definiujemy tak, jak w poprzednim przykładzie, zaznaczając w zakładce „Stałe…" opcję „Uwzględnij wyraz wolny" i zostawiając puste pole „Model". Podobnie w zakładce „Losowe…", pole „Model" zostawiamy puste, zaznaczamy jedynie opcję „Uwzględnij wyraz wolny" i dodajemy zmienną *OSOBA* do pola „Kombinacje". Wykonujemy analizę.

Parametry wszystkich szacowanych w tym przykładzie modeli prezentuje Tabela 6. Wartość stałej w Modelu 0 opisuje nam średnią ogólną (wyniki wszystkich osób we wszystkich pomiarach), co w tym wypadku nie stanowi informacji przydatnej do interpretowania wyników. Wariancja wyników została rozbita na dwie składowe: międzyosobową (efekt losowy dla stałej równania regresji) i wewnątrzosobową (reszta). Dzieląc wariancję międzyosobową przez wariancję całkowitą, otrzymujemy wskaźnik tzw. korelacji wewnątrzklasowej (*intraclass correlation*, ICC). W tym wypadku wynosi on:

$$\mathrm{ICC} = \frac{20,495}{20,495 + 136,895} = 0,13$$

Współczynnik ten oznacza, że około 13% całkowitej wariancji wyników jest związane ze zmiennością międzyosobową. Korelację wewnątrzklasową można też interpretować jako oczekiwaną korelację między dwoma losowo wybieranymi pomiarami u tej samej osoby.

W następnym kroku dodajemy do modelu efekty stałe. W przypadku badań, w których głównym predyktorem jest czas, w tym kroku należy skupić się na określeniu linii trendu najlepiej dopasowanej do danych. W niniejszym przykładzie etap ten uprościmy, na podstawie wykresu

*Tabela 6.*
Oszacowanie parametrów z hierarchicznego modelu liniowego – przykład 2

| Parametr | Model 0 | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|---|
| | | | Efekty stałe: | | |
| Stała | 15,71*** | 37,08*** | 37,56*** | 40,37*** | 41,89*** |
| *Czynniki pierwszego poziomu* | | | | | |
| POMIAR | | –9,32*** | –9,32 *** | –9,32*** | –10,83*** |
| POMIAR2 | | 0,72*** | 0,72*** | 0,72*** | 0,81*** |
| *Czynniki drugiego poziomu* | | | | | |
| SSZ | | | –0,95 | –6,58 | –9,63 |
| *Interakcje międzypoziomowe* | | | | | |
| POMIAR · SSZ | | | | | 3,01* |
| POMIAR2 · SSZ | | | | | –0,17 |
| | | | Parametry losowe: | | |
| Reszta | 136,895 | 33,204 | 33,204 | 19,381 | 19,100 |
| Stała (wariancja) | 20,495 | 30,864 | 30,638 | 85,701 | 79,226 |
| POMIAR (wariancja) | | | | 1,314 | 0,770 |
| –2 · LL | 941,833 | 788,846 | 788,767 | 765,176 | 757,595 |

*Uwaga*. Model 0 to model ze stałą równania regresji. Model 1 to model z efektem stałym dla trendu czasowego. Model 2 to model z efektem stałym dla trendu czasowego i efektem międzyosobowym. Model 3 to model z dodanym efektem losowym dla trendu czasowego. Model 4 to model z efektem losowym dla trendu czasowego i interakcją międzypoziomową. Jako metodę estymacji zastosowano metodę najwyższej wiarygodności (ML).

* $p < 0{,}05$; *** $p < 0{,}001$

od razu zakładając, że najlepiej dopasowana do danych jest funkcja kwadratowa czasu pomiaru. Aby zawrzeć taką funkcję w naszym modelu, uzupełniamy zakładkę „Stałe…", w polu „Model" umieszczając zmienne *POMIAR* i *POMIAR2*, bez interakcji między nimi (alternatywnie, możemy w polu „Model" umieścić zmienną *POMIAR* i interakcję *POMIAR · POMIAR*, co pozwoli nam uzyskać identyczny efekt). Wykonujemy analizę. Jak widać w Tabeli 6, w otrzymanym w ten sposób Modelu 1 w znaczący sposób zmniejszyła się wewnątrzosobowa wariancja błędu. Siła efektu, obliczona zgodnie z wzorem (1), wynosi:

$$R_1^2 = \frac{136{,}895 - 33{,}204}{136{,}895} = 0{,}76$$

co oznacza, że model, w którym wynik w teście *SMK* ma charakter funkcji kwadratowej czasu pomiaru, tłumaczy około trzech czwartych wewnątrzosobowej zmienności wyników. Warto zauważyć, że wariancja międzyosobowa związana ze stałą równania regresji wzrosła w stosunku do modelu poprzedniego. Próbując w sposób analogiczny do $R^2{}_1$ obliczyć wartość $R^2$ dla poziomu drugiego (międzyosobowego), otrzymalibyśmy wartość ujemną, której nie da się interpretować w kategoriach proporcji wyjaśnionej wariancji – jest to dobry przykład problemów z obliczaniem siły efektu w analizach wielopoziomowych.

W kolejnym kroku powinniśmy włączyć do modelu zmienne drugiego poziomu, by sprawdzić, czy tłumaczą one wariancję międzyosobową. Uzupełniamy więc pole „Model" w zakładce „Stałe…" o zmienną *SSZ*, bez interakcji z pozostałymi zmiennymi. Wykonujemy analizę. W otrzymanym w ten sposób Modelu 2 (por. Tabela 6) widzimy, że parametr dla zmiennej *SSZ* nie jest istotny statystycznie, a zmiana wariancji międzyosobowej w stosunku do Modelu 1 jest minimalna (wynosi zaledwie 0,226). Oznacza to, że Styl Skoncentrowany na Zadaniu nie tłumaczy międzyosobowego zróżnicowania wyników. Warto też zwrócić uwagę, że wariancja resztowa jest identyczna jak w Modelu 1, co jest efektem oczekiwanym, gdyż zmienna międzyosobowa nie powinna wpływać na wariancję wewnątrzosobową.

Rysunek 4 sugeruje, że osoby badane różnią się nie tylko poziomem wykonania, ale też przebiegiem czasowym (u jednych osób spadek wykonania jest dosyć gwałtowny, u innych przebiega to łagodniej). Naszym kolejnym celem jest więc sprawdzenie, czy to zróżnicowanie jest istotne oraz czy istnieje interakcja między przebiegiem czasowym a Stylem Skoncentrowanym na Zadaniu. Aby to sprawdzić, musimy najpierw włączyć do analizy hierarchicznej efekty losowe dla zmiennych opisujących czas pomiaru, dopuszczając w ten sposób oszacowanie parametrów nachylenia linii regresji oddzielnie dla każdej osoby badanej i oszacowanie wariancji grupowych współczynników. Tak skonstruowany model będzie następnie modelem bazowym do testowania interakcji.

Aby włączyć do modelu efekty losowe dla zmiennych pierwszego poziomu, uzupełniamy zakładkę „Losowe…", wstawiając w pole „Model" zmienne *POMIAR* i *POMIAR2*, bez interakcji (podobnie jak robiliśmy to w zakładce „Stałe…"). Estymacja modelu z efektami losowymi dla trendu liniowego i kwadratowego nie przynosi pozytywnych rezultatów (ostrzeżenie na początku raportu SPSS wynika z faktu, że dla składnika losowego zmiennej *POMIAR2* algorytm oszacował wartość 0). Dzieje się tak często, gdy w modelu zamieścimy zbyt wiele zbędnych efektów losowych. Bez problemu za to zostaje oszacowany model z efektem losowym tylko dla trendu liniowego (w zakładce „Losowe…" w polu „Model" zostawiamy jedynie zmienną *POMIAR*). Parametry tego modelu (Model 3) prezentuje Tabela 6. Model ten oznacza, że badani różnią się zarówno początkowym poziomem wykonania testu, jak i szybkością spadku (ujemna wartość parametru efektu stałego dla zmiennej *POMIAR*), nie różnią się natomiast zmianą tej szybkości (u każdego badanego ma ona stałą, zmniejszającą się wartość, co odzwierciedla parametr efektu *POMIAR2* o przeciwnym znaku). Ujmując to inaczej – kształt przebiegu czasowego jest w tym modelu identyczny u każdego z badanych, różnią się oni jednak położeniem punktu minimalnego prezentującej ten przebieg czasowy paraboli. Przeszacowaniu uległy parametry wariancji wewnątrzosobowej i międzyosobowej.

Skoro nie zawsze możemy opierać się na wartościach $R^2$, powstaje pytanie, jak ocenić, czy kolejny model jest lepszy od poprzedniego, tj. czy w sposób istotny lepiej opisuje prawdziwy układ wyników. Służą do tego tzw. kryteria informacyjne charakteryzujące szacowany model – podstawowym kryterium jest logarytmiczny wskaźnik wiarygodności (*log-likelihood*, LL), pozostałe (najczęściej wykorzystywane to AIC lub BIC, choć raport SPSS oferuje większą ich liczbę) są obliczane w oparciu o tę wartość, przy uwzględnieniu korekty związanej z liczbą parametrów modelu (im więcej parametrów, tym bardziej obciążona zostaje wartość kryterium, analogicznie do skorygowanego $R^2$ w klasycznej regresji liniowej). Podstawowa reguła jest prosta – im bliższa zera jest wartość kryterium, tym model jest lepiej dopasowany.

Gdy potrzebujemy większej precyzji w ocenie, można obliczyć, czy różnica w wielkości kryterium jest dla dwóch porównywanych modeli istotna statystycznie (pod warunkiem, że jeden z tych modeli jest zagnieżdżony w drugim). SPSS nie pozwala na automatyczne porównanie, trzeba więc zrobić to ręcznie – różnica między dwiema wartościami $-2 \cdot$ LL ma rozkład $\chi^2$ o liczbie stopni swobody równej różnicy w liczbie parametrów porównywanych modeli. Porównywać w ten sposób można jedynie modele oszacowane metodą największej wiarygodności (*maximum likelihood*, ML). W przypadku modeli szacowanych metodą warunkowej największej wiarygodności (*restricted maximum likelihood*, REML) można porównywać w ten sposób modele różniące się jedynie liczbą efektów losowych, przy zachowaniu tych samych efektów stałych.

W analizowanym przypadku, Model 3 jest w sposób istotny ($p < 0{,}001$) lepiej dopasowany do danych od Modelu 2. Jak widać, oszacowania parametrów wewnątrzosobowych nie zmieniły się, jednak liniowy trend związany z czasem pomiaru przybrał charakter efektu losowego, więc wartość –9,32 traktujemy jedynie jako wartość oczekiwaną, odzwierciedlającą szybkość spadku, która to wartość charakteryzuje się w badanej populacji zmiennością opisaną przez składnik wariancji zawarty w modelu. Jak widać w Tabeli 6, wariancja dla tego efektu wynosi 1,314, co oznacza, że odchylenie standardowe wynosi w przybliżeniu 1,15. Ponieważ 95% wyników leży w obrębie około dwóch odchyleń standardowych (dokładna wartość to 1,96) możemy wyliczyć, że:

$$-9{,}32 + 1{,}96 \cdot 1{,}15 = -7{,}07$$
$$-9{,}32 - 1{,}96 \cdot 1{,}14 = -11{,}57$$

co oznacza, że współczynnik opisujący liniową szybkość spadku poziomu wykonania waha się (z prawdopodobieństwem 95%) od 7 do prawie 12 – mniej więcej o tyle procent czasu w obszarze idealnym zmniejsza się poziom wykonania w pierwszej minucie testu. Obie granice przedziału są w tym wypadku ujemne, nie ma więc praktycznie osób, u których liniowy trend czasowy byłby rosnący. Zdarza się to jednak w niektórych badaniach i ciekawe jest wtedy wyliczenie, u jakiego procenta osób trend czasowy wskazuje na wzrost, a u jakiego – na spadek wyników.

Oczywiście, ze względu na składnik kwadratowy trendu (który może być interpretowany jako przyspieszenie), szybkość spadku zmienia się w każdej kolejnej minucie testu, współczynnik przypisany parametrowi *POMIAR* nie ma więc stałej i jednoznacznej interpretacji dla całego przebiegu badania. Spadek poziomu wykonania jest coraz wolniejszy, aż do osiągnięcia minimum, po którym trend zmienia kierunek i staje się rosnący. Punkt minimum dla trendu kwadratowego można w tym wypadku obliczyć wzorem

$$(2) \qquad CZAS_{\text{minimum}} = \frac{-T}{2 \cdot T2}$$

gdzie *T* i *T2* to współczynniki regresji odpowiednio dla liniowego i kwadratowego składnika trendu. Podstawiając za *T* wartości opisujące przedział ufności dla współczynnika liniowego, a za *T2* wartość 0,72, będącą wartością efektu stałego dla zmiennej *POMIAR2*, otrzymujemy wartości 4,92 i 8,02. Pamiętając, że pierwsza minuta została zakodowana jako wartość zero, możemy stwierdzić, że z 95-procentową pewnością poziom wykonania testu spada do wartości minimalnej (po czym zaczyna rosnąć) między szóstą a dziewiątą minutą trwania zadania.

W ostatnim kroku wprowadzamy do modelu interakcje międzypoziomowe, by oszacować, w jakim stopniu Styl Skoncentrowany na Zadaniu odpowiada za przebieg zmian w poziomie wykonania testu. W SPSS w zakładce „Stałe…” uzupełniamy pole „Model” o interakcję *POMIAR · SSZ* oraz *POMIAR2 · SSZ*. Wykonujemy analizę.

Model 4 wykazuje lepsze dopasowanie od Modelu 3 na poziomie istotnym statystycznie ($p < 0{,}05$), jednak istotna okazała się jedynie interakcja *POMIAR · SSZ*. Interakcja z kwadratowym składnikiem trendu nie osiągnęła istotności statystycznej, co oznacza, że poziom SSZ ma związek z szybkością zmiany poziomu wykonania, nie ma jednak wpływu na kształt (tj. zakrzywienie) linii trendu. Jak widać, wariancja wewnątrzosobowa (reszta) jest praktycznie identyczna jak w Modelu 3, obniżeniu uległa natomiast wariancja międzyosobowa. Dla efektów losowych Modelu 4 modelem bazowym jest Model 3 i w oparciu o jego parametry można spróbować obliczyć przybliżone $\Delta R^2$ określające, w jakim stopniu wariancja w nachyleniu linii regresji jest tłumaczona przez zmienną międzyosobową. Posługując się równaniem analogicznym do równania (1), podstawiając jedynie za resztę interesujący nas składnik wariancji, możemy obliczyć, że:

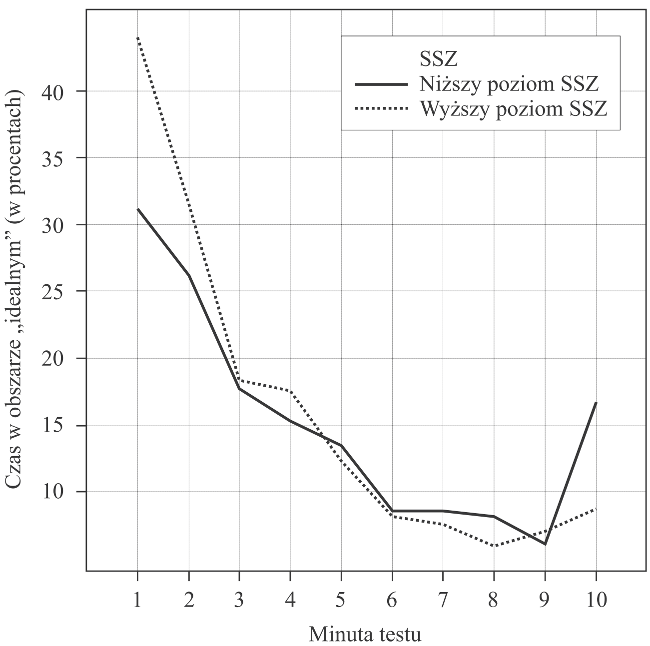
$$R^2_{\text{nachylenie}} = \frac{1{,}314 - 0{,}77}{1{,}314} = 0{,}41$$

Oznacza to, że Styl Skoncentrowany na Zadaniu tłumaczy około 41% zmienności w składniku linowym zmiany poziomu wykonania podczas całego testu *SMK*. $R^2_1$ dla Modelu 4 wynosi zaś:

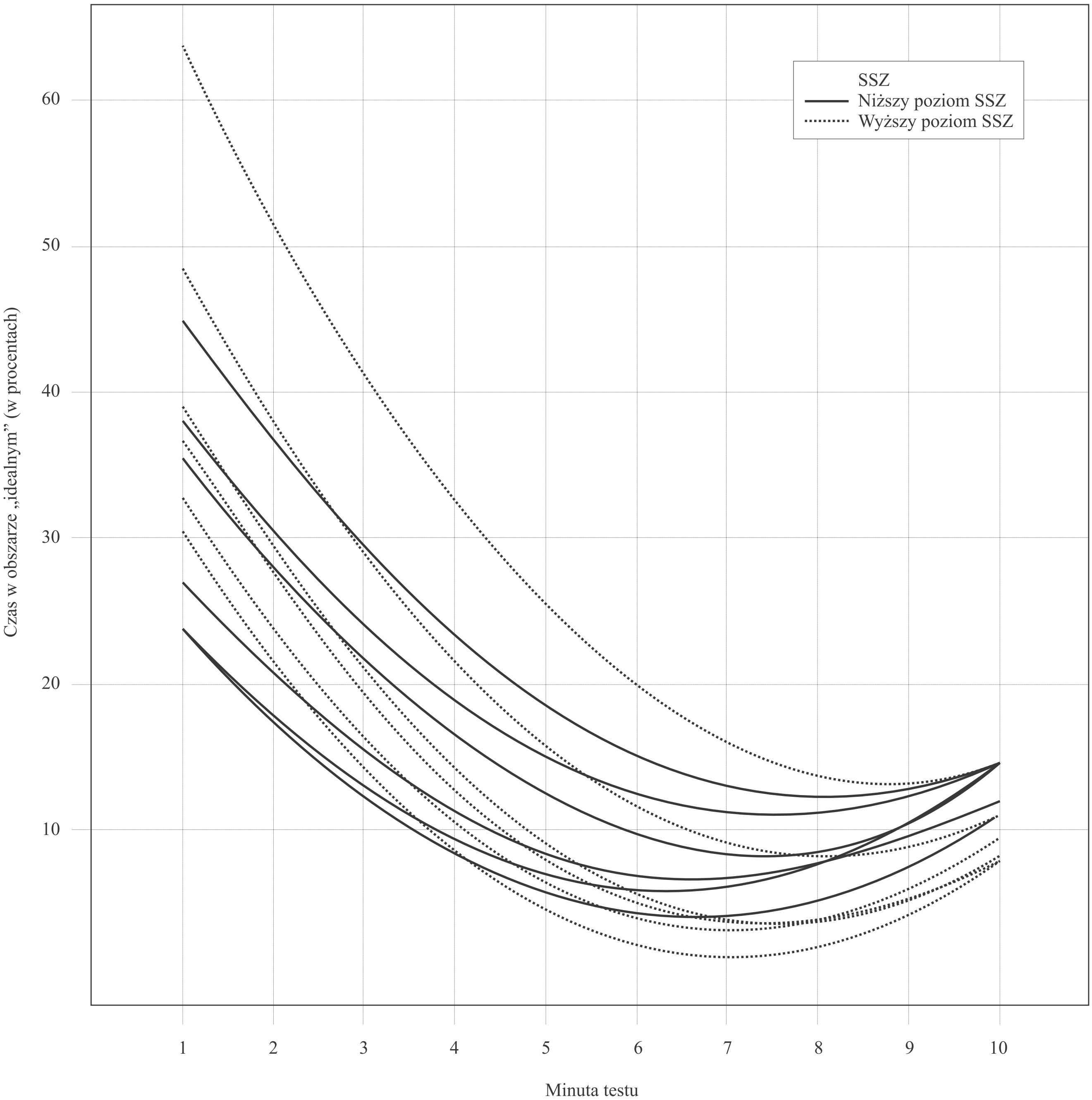
$$R^2_1 = \frac{136{,}895 - 19{,}1}{136{,}895} = 0{,}86$$

Przewidywane indywidualne linie trendu dla poszczególnych osób badanych zgodnie z Modelem 4 prezentuje Rysunek 6. Widać na nim wyraźnie, że poziom wykonania testu *SMK* zmniejsza się wraz z upływem czasu (można to przypisać efektowi zmęczenia), aż do osiągnięcia punktu krytycznego około siódmej minuty, po której dodatni wektor przyspieszenia (pozytywny współczynnik dla kwadratu czasu pomiaru) powoduje stabilizację wyników (można to przypisać efektowi nabierania wprawy), a u niektórych osób nawet nieznaczny ich wzrost. Początkowy poziom wykonania nie jest istotnie związany z poziomem Stylu Skoncentrowanego na Zadaniu, ale u osób z większym natężeniem SSZ obniżenie wykonania jest większe, tj. nie jest niwelowane w tak dużym stopniu przez efekt wprawy, co jest wyraźne już od około trzeciej minuty (może to wynikać np. z faktu, że determinacja tych osób, by wypaść dobrze, wymaga więcej wysiłku, co z kolei zwiększa efekt zmęczenia i powoduje szybszy spadek poziomu wykonania). Ostateczny model wyjaśnia około 86% zróżnicowania wewnątrzosobowego, podczas gdy poziom SSZ odpowiada za około 41% międzyosobowej zmienności w szybkości zmiany poziomu wykonania.

Powyższy przykład dobrze ilustruje różnicę między testowaniem hipotez w modelu ANOVA i w hierarchicz-




*Rysunek 5.*
Przeciętny poziom wykonania testu SMK u grup różniących się poziomem SSZ (przykład drugi).




*Rysunek 6.*
Przebiegi czasowe dla poszczególnych osób – wartości przewidywane zgodnie z Modelem 4 (przykład drugi).

nym modelu liniowym. W analizie wariancji z czynnikiem wewnątrzosobowym testowaliśmy hipotezę, że trend czasowy ma np. charakter trendu kwadratowego i odpowiada za zróżnicowanie wyników w kolejnych pomiarach w takim samym stopniu u wszystkich osób badanych. Wprowadzając dwukategorialną zmienną międzyosobową SSZ i interakcję z czasem pomiaru, testowaliśmy hipotezę, że w wynikach mamy dwa trendy czasowe, odmienne w każdej z grup, ale za to identyczne wewnątrzgrupowo. Gdybyśmy konserwatywnie, opierając się na analizie z poprawką Greenhouse–Geissera, uznali brak istotności statystycznej dla tej interakcji, bylibyśmy zmuszeni odrzucić taką hipotezę i przyjąć, że zmiana wyników w czasie ma dla każdej z osób identyczny charakter, niezależny od poziomu SSZ, a różnice w wynikach mają jedynie charakter błędu pomiaru.

W hierarchicznym modelu liniowym tok rozumowania jest nieco inny. Po pierwsze, testujemy hipotezę, że zmiana wyników w czasie ma charakter trendu kwadratowego (Model 1), co jest spójne z modelem ANOVA bez czynnika międzyosobowego. W Modelu 2 sprawdzamy, czy ogólny poziom wykonania (ale nie kształt linii trendu) ma związek z wartością SSZ – ten krok jest analogiczny do modelu ANOVA z czynnikiem międzyosobowym, bez interakcji czynnika wewnątrz- i międzyosobowego. Taki model w analizie wariancji praktycznie nie bywa analizowany, domyślnie zwykle testujemy model pełny, tj. z interakcją. Dalej różnice są wyraźniejsze. Dodając efekt losowy dla trendu czasowego w Modelu 3, dopuszczamy indywidualne zróżnicowanie zmian wyników w czasie i staramy się to jak najlepiej opisać za pomocą oszacowanych parametrów. W ostatnim kroku, dodając interakcję dwukategorialnej zmiennej międzyosobowej z wewnątrzosobowym trendem czasowym, sprawdzamy, w jakim stopniu czynnik międzyosobowy jest w stanie wytłumaczyć to zróżnicowanie indywidualne. Poziom SSZ służy więc jedynie do wyjaśniania tak czy inaczej zaobserwowanego efektu, a nie – jak w analizie wariancji – do określenia, czy osoby badane różnią się pod względem tego, jak zmienia się w czasie ich poziom wykonania zadania.

## Podsumowanie

W niniejszym artykule zaprezentowane zostały dwa podejścia do analizy danych ze schematu badawczego z powtarzanymi pomiarami. Klasyczna i do tej pory najpopularniejsza analiza tzw. modelu mieszanego, czyli analiza wariancji z powtarzanym pomiarem, została zestawiona z hierarchicznym modelem liniowym (zwanym też liniowym modelem mieszanym lub wielopoziomową analizą regresji), w której podobny cel realizowany jest metodami największej wiarygodności. W przypadku, gdy zebrane dane spełniają założenia niezbędne do przeprowadzenia trafnej analizy wariancji, ANOVA dla powtarzanych pomiarów wydaje się intuicyjnym i godnym polecenia wyborem, zapewniając łatwą interpretację uzyskanych wyników. Jeśli jednak założenia te (dotyczące m.in. sferyczności) nie są spełnione, lub gdy struktura zmiennych jest zbyt złożona do ujęcia w czynnikową analizę wariancji, najlepszym rozwiązaniem staje się hierarchiczny model liniowy. Analiza ta wymaga co prawda znacznie większej mocy obliczeniowej, jednak przy możliwościach współczesnych komputerów argument ten jest praktycznie nieistotny. Analiza wielopoziomowa wymaga też od badacza zrozumienia struktury danych na dużo głębszym poziomie (wprowadzanie czynników losowych, interakcje między poziomami hierarchii itp.) niż w przypadku analizy wariancji, co jest niezbędne do stworzenia i interpretacji trafnego modelu. Jest to oczywiście tylko pozorna wada, gdyż w praktyce ten dodatkowy wysiłek prowadzi do lepszego zrozumienia opisywanej rzeczywistości.

Współczesne podręczniki poświęcone analizie danych z powtarzanych pomiarów coraz częściej sugerują liniowy model mieszany jako podstawową i najlepszą metodę statystyczną (m.in. Davis, 2002; Hedeker i Gibbons, 2006; Verbeke i Molenberghs, 2000), czasem posuwając się wręcz do traktowania jednozmiennowej analizy wariancji jako metody przestarzałej i mającej zbyt wiele ograniczeń. Zwiększająca się popularność analizy wielopoziomowej widoczna jest też na gruncie psychologii – choć jeszcze piąta edycja *APA Publication Manual* (American Psychological Association, 2001) nie zawierała wzmianki na temat sposobów prezentacji wyników z analizy wielopoziomowej, najnowsze, szóste wydanie (2010) zawiera już przykładowe tabele wskazujące sposób przedstawiania parametrów estymowanych modeli (zrezygnowano natomiast w tym wydaniu z przykładu tabeli dla analizy wariancji). Można więc przewidywać, że znajomość tej metody w niedługim czasie będzie nie tylko przydatnym, lecz także niezbędnym elementem warsztatu badawczego.

## LITERATURA CYTOWANA


American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (wyd. 5). Washington, DC: APA.

American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (wyd. 6). Washington, DC: APA.

Bates, D. M., Maechler, M. (2009). *Lme4: Linear mixed-effects models using S4 classes*. R package. http://CRAN.R-project.org/package=lme4.

Bedyńska, S., Brzezicka, A. (red.) (2007). *Statystyczny drogowskaz. Praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii*. Warszawa: Wydawnictwo SWPS Academica.

Brzeziński, J. (1996). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.

Brzeziński, J. (2008). *Badania eksperymentalne w psychologii i pedagogice* (wyd. 2). Warszawa: Wydawnictwo Naukowe Scholar.

Brzeziński, J., Stachowski, R. (1984). *Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych* (wyd. 2). Warszawa: Państwowe Wydawnictwo Naukowe.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements.* Heidelberg: Springer-Verlag.

Ferguson, G. A., Takane, Y. (1997). *Analiza statystyczna w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.

Fidell, L. S., Tabachnick, B. G. (2003). Preparatory data analysis. W: J. A. Schinka, W. F. Velicer, I. B. Weiner (red.). *Handbook of psychology. Volume 2: Research methods in psychology* (s. 115–141). Hoboken, New Jersey: John Wiley & Sons.

Goldstein, H. (1999). *Multilevel statistical models*. London: University of London, Institute of Education.

Hedeker, D., Gibbons, R. D. (2006). *Longitudinal data analysis*. New Jersey: John Wiley & Sons.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT--Publikaties.

Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Olejnik, S., Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434–447.

Pinheiro, J. C., Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.

Pinheiro, J. C., Bates, D. M., DebRoy, S., Sarkar, D., R Core team (2009). *Nlme: Linear and nonlinear mixed effects models*. R package. http://CRAN.R-project.org/package=nlme.

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (wyd. 2). Thousand Oaks: Sage Publications.

Shaughnessy, J. J., Zechmeister, E. B., Zechmeister, J. S. (2002). *Metody badawcze w psychologii*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Snijders, T. A. B., Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

Strelau, J., Jaworowska, A., Wrześniewski, K., Szczepaniak, P. (2005) *Kwestionariusz Radzenia Sobie w Sytuacjach Stresowych CISS. Podręcznik.* Warszawa: Pracownia Testów Psychologicznych PTP.

Verbeke, G., Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

## Załącznik A

| OSOBA | AK_50 | AK_70 | AK_S |
|---|---|---|---|
| 1 | 803 | 765 | 922 |
| 2 | 720 | 727 | 880 |
| 3 | 838 | 768 | 908 |
| 4 | 765 | 716 | 844 |
| 5 | 883 | 787 | 954 |
| 6 | 818 | 725 | 916 |
| 7 | 798 | 729 | 988 |
| 8 | 800 | 747 | 913 |
| 9 | 820 | 755 | 860 |
| 10 | 916 | 788 | 986 |
| 11 | 812 | 713 | 909 |
| 12 | 724 | 657 | 827 |
| 13 | 768 | 714 | 909 |
| 14 | 841 | 712 | 917 |
| 15 | 785 | 729 | 926 |
| 16 | 728 | 698 | 847 |
| 17 | 785 | 788 | 878 |
| 18 | 740 | 746 | 856 |
| 19 | 826 | 758 | 951 |
| 20 | 755 | 709 | 843 |

*Uwaga*. Tabela zawiera dane do pierwszego z analizowanych w tekście przykładów. Zastosowano szeroki format danych. Kolumna AK_50 zawiera przeciętny czas reakcji w sytuacji zadania wykonywanego w tempie 50 bodźców na minutę, kolumna AK_70 zawiera wynik z tempa 70 bodźców na minutę, a kolumna AK_S wyniki z tempa swobodnego.

**Załącznik B**

| OSOBA | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | SSZ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 24 | 14 | 10 | 9 | 13 | 6 | 2 | 1 | 1 | 12 | N |
| 2 | 41 | 30 | 21 | 14 | 12 | 8 | 1 | 9 | 4 | 14 | N |
| 3 | 38 | 31 | 26 | 19 | 14 | 12 | 12 | 11 | 6 | 21 | N |
| 4 | 42 | 45 | 29 | 27 | 16 | 11 | 12 | 10 | 9 | 18 | N |
| 5 | 16 | 22 | 11 | 6 | 12 | 9 | 11 | 6 | 11 | 17 | N |
| 6 | 27 | 15 | 10 | 17 | 14 | 5 | 13 | 12 | 6 | 19 | N |
| 7 | 42 | 22 | 20 | 10 | 6 | 6 | 8 | 3 | 2 | 6 | W |
| 8 | 36 | 19 | 8 | 8 | 12 | 7 | 10 | 3 | 7 | 7 | W |
| 9 | 50 | 45 | 25 | 19 | 16 | 11 | 3 | 6 | 13 | 15 | W |
| 10 | 61 | 56 | 37 | 45 | 23 | 15 | 17 | 13 | 14 | 20 | W |
| 11 | 34 | 19 | 6 | 5 | 9 | 6 | 2 | 6 | 4 | 2 | W |
| 12 | 42 | 30 | 14 | 18 | 8 | 4 | 6 | 5 | 3 | 2 | W |

*Uwaga*. Tabela zawiera dane do drugiego z analizowanych w tekście przykładów. Zastosowano szeroki format danych. Kolumny od P1 do P10 zawierają wyniki z kolejnych minut wykonywania testu. Kolumna SSZ zawiera dane na temat poziomu cechy Styl Skoncentrowany na Zadaniu (N oznacza wynik niższy, zaś W wyższy).

**Załącznik C**

Zbiory danych i polecenia do programu *R* odtwarzające przykładowe analizy wielopoziomowe prezentowane w artykule. Całość można przepisać lub (w przypadku elektronicznej kopii tekstu) przekopiować do konsoli *R*. Wiersze poprzedzone # automatycznie zostaną pominięte.

Ze względu na stosowane algorytmy, wartości parametrów modeli oszacowanych w środowisku *R* mogą być nieco inne niż otrzymane w pakiecie SPSS i opisane w artykule. Różnice między tymi programami nie powinny jednak nigdy być na tyle znaczące, by zmieniać interpretację wyników.

```
# Inicjowanie modułu do analizy
library(nlme)
# Ustawienie kodowania zmiennych kategorialnych na zgodne
# z pakietem SPSS
options(contrasts=c(„contr.SAS”, „contr.poly”))
# inne kodowania uzyskamy, zastępując wyrażenie contr.SAS
# nazwą innego kontrastu (opis definiowania kontrastów w
# pakiecie R uzyskamy po wpisaniu w konsoli polecenia
# ? contrasts)
#
# PRZYKŁAD PIERWSZY
#
# Zmienna CZAS REAKCJI w formacie długim
RT <- c(803, 765, 922, 720, 727, 880, 838, 768, 908, 765, 716, 844, 883, 787,
954, 818, 725, 916, 798, 729, 988, 800, 747, 913, 820, 755, 860, 916, 788,
986, 812, 713, 909, 724, 657, 827, 768, 714, 909, 841, 712, 917, 785, 729,
926, 728, 698, 847, 785, 788, 878, 740, 746, 856, 826, 758, 951, 755, 709,
843)
# Tworzenie tabeli wyników w formacie długim
ID <- factor(rep(1:20, each=3))
TYPE <- factor(rep(c(„AK_50”, „AK_70”, „AK_S”), times=20))
```

```
BAZA1 <- data.frame(ID, TYPE, RT)
names(BAZA1) <- c("OSOBA", "TYP_AK", "CZAS_R")
# Estymacja modelu opisanego w tekście artykułu
MA1 <- lme(CZAS_R~1+TYP_AK, random=~1|OSOBA, data=BAZA1)
# Tworzenie tzw. modelu zerowego dla modelu MA1
MA0 <- update(MA1, CZAS_R~1)
#
# PRZYKŁAD DRUGI
#
# Zmienna CZAS W OBSZARZE IDEALNYM w formacie długim
IDEAL <- c(24, 14, 10, 9, 13, 6, 2, 1, 1, 12, 41, 30, 21, 14, 12, 8, 1, 9, 4,
14, 38, 31, 26, 19, 14, 12, 12, 11, 6, 21, 42, 45, 29, 27, 16, 11, 12, 10, 9,
18, 16, 22, 11, 6, 12, 9, 11, 6, 11, 17, 27, 15, 10, 17, 14, 5, 13, 12, 6, 19,
42, 22, 20, 10, 6, 6, 8, 3, 2, 6, 36, 19, 8, 8, 12, 7, 10, 3, 7, 7, 50, 45,
25, 19, 16, 11, 3, 6, 13, 15, 61, 56, 37, 45, 23, 15, 17, 13, 14, 20, 34, 19,
6, 5, 9, 6, 2, 6, 4, 2, 42, 30, 14, 18, 8, 4, 6, 5, 3, 2)
# Tworzenie tabeli wyników w formacie długim
ID <- factor(rep(1:12, each=10))
TIME <- rep(0:9, times=12)
TIME2 <- TIME^2
SSZ <- factor(rep(c(„N_SSZ", „W_SSZ"), each=60))
BAZA3 <- data.frame(ID, TIME, TIME2, IDEAL, SSZ)
names(BAZA3) <- c("OSOBA", "POMIAR", "POMIAR2", "WYNIK", "SSZ")
# Model zerowy
MB0 <- lme(WYNIK~1, random=~1|OSOBA, method="ML", data=BAZA3)
# Model 1
MB1 <- lme(WYNIK~1+POMIAR+POMIAR2, random=~1|OSOBA, method="ML", data=BAZA3)
# Model 2
MB2 <- lme(WYNIK~1+POMIAR+POMIAR2+SSZ, random=~1|OSOBA, method="ML",
data=BAZA3)
# Model 3 – ten model nie osiąga zbieżności
# (nie daje się oszacować)
MB3 <- lme(WYNIK~1+POMIAR+POMIAR2+SSZ, random=~1+POMIAR+POMIAR2|OSOBA,
method="ML", data=BAZA3)
# Model 3 (właściwy)
MB3 <- lme(WYNIK~1+POMIAR+POMIAR2+SSZ, random=~1+POMIAR|OSOBA, method="ML",
data=BAZA3)
# Model 4
MB4 <- lme(WYNIK~1+(POMIAR+POMIAR2)*SSZ, random=~1+POMIAR|OSOBA, method="ML",
data=BAZA3)
#
# Nazwa kodowa każdego z szacowanych modeli znajduje się
# przed znakiem „<-" (np. MA0, MA1, MB1, MB4)
#
# Dla każdego z modeli można m.in.:
# – wyświetlić parametry poleceniem:
# summary(Nazwa.modelu), np.:
summary(MB4)
# – wykonać omnibusowy test F poleceniem
# anova(Nazwa.modelu, type="marginal"), np.:
anova(MB4, type="marginal")
```

```
# – wyświetlić składniki wariancji poleceniem:
# VarCorr(Nazwa.modelu), np.:
VarCorr(MA1)
#
# Porównać dopasowanie kilku modeli można poleceniem:
# anova(Nazwa.modelu 1, Nazwa.modelu 2,…), np.:
anova(MB0, MB1, MB2, MB3)
```

# Multilevel analysis for repeated measures – hierarchical linear model as an alternative to the analysis of variance

Piotr Zieliński

*Military Institute of Aviation Medicine, Warsaw*

**Abstract**

The repeated measures design is a popular research method, in some cases more suitable then the independent groups design. The most common method of data analysis for the repeated measures is the mixed effects univariate analysis of variance (ANOVA). The strict assumptions of this statistical method, however, cannot always be met, therefore an alternative is needed. This article presents the analysis of the repeated measures design in the hierarchical linear model, which is characterized by more liberal assumptions and better modelling ability. In two examples the author describes how in a simple research design the results of ANOVA and the hierarchical linear model are almost identical, while in more complex designs the hierarchical linear model allows for more accurate analyses than ANOVA.

*Key words:* repeated measures design, hierarchical linear model, repeated measures ANOVA

Złożono: 5.05.2010
Złożono poprawiony tekst: 7.08.2010
Zaakceptowano do druku: 28.08.2010