

Wielowymiarowa analiza wariancji – MANOVA

Elżbieta Aranowska¹, Jolanta Rytel²

¹ Wydział Psychologii, Uniwersytet Warszawski

² Instytut Psychologii, Uniwersytet Kardynała Stefana Wyszyńskiego

Artykuł dotyczy modelu wielowymiarowej analizy wariancji (MANOVA). W ramach wprowadzenia przedstawiono różnice między tą metodą a jednowymiarową analizą wariancji (ANOVA), równocześnie – w ramach opisu podstawowych planów badawczych z powtarzaniem pomiarów na tej samej populacji – pokazano te schematy badawcze, które dostarczają takich danych, które z kolei mogą być analizowane wyłącznie metodami MANOVA. Opisując strukturę formalną modelu, zaprezentowano podstawowe definicje z nim związane, odwołując się do odpowiadających im pojęć ANOVA i do naturalnego, intuicyjnego ich rozszerzenia w MANOVA, wykorzystując wcześniejszą dyskusję o niezależności wartości oczekiwanych zmiennych i wartości miar związku dla par zmiennych (a dokładniej – niezależności średnich arytmetycznych i wartości współczynnika korelacji r -Pearsona). Zaprezentowano, na czym polega rozszerzenie założeń MANOVA, postaci hipotez zerowych oraz statystyk testu. Zwrócono uwagę na niejednoznaczność rozwiązania formalnego (brak jednego ustalonego sprawdzianu testu) i przedstawiono te statystyki, które najczęściej pojawiały się w pakietach statystycznych ostatnich dwu dziesięcioleci. Ilustracją dla przedstawionych rozwiązań formalnych był fikcyjny przykład dobrany dla najprostszego planu jednoczynnikowej, dwuwymiarowej analizy wariancji, dla którego wyznaczono zarówno ręcznie, jak i za pomocą pakietu SPSS wartości wszystkich wprowadzanych statystyk. Przedstawiono także przykład aplikacji wielowymiarowej analizy wariancji w badaniach psychologicznych dotyczących oceny efektywności pracy menedżerów, równocześnie podkreślając niezbędność komplementarnego stosowania dwu statystycznych metod analizy danych: wielowymiarowej analizy wariancji i analizy dyskryminacyjnej.

Słowa kluczowe: wielowymiarowa analiza wariancji MANOVA, analiza dyskryminacyjna, metody wielowymiarowe, statystyczne modele analizy danych

Plany badawcze dla metod analizy wariancji

Metodologia statystyki, zależna od rozumowania indukcyjnego, nie jest w pełni skodyfikowana ani też wolna od kontrowersji. Różni użytkownicy, analizując ten sam zbiór danych, mogą dochodzić do odmiennych wniosków. Zazwyczaj dane zawierają więcej informacji niż można ujawnić za pomocą dostępnych narzędzi statystycznych. To, w jakim stopniu użytkownikowi uda się

wydobyc te informacje, zależy nie tylko od jego wiedzy, ale również od wprawy i doświadczenia. Czyni to statystykę sztuką polegającą na dokonywaniu trafnych wyborów (Rao, 1994). Niełatwo dokonywać takich wyborów bez rzetelnej wiedzy na temat podstaw wybranych metod statystycznych oraz kryteriów wyboru odpowiedniej metody analizy. Niełatwo także wykorzystywać te metody w kompetentny sposób bez znajomości ich ograniczeń oraz uprawnionych (bądź nie) interpretacji uzyskanych rezultatów analizy. Dynamicznie narastająca transmisja osiągnięć współczesnych nauk społecznych, a także innych dyscyplin empirycznych stawia przed osobami je uprawiającymi (zarówno w sposób czynny, jak i bierny) coraz trudniejsze do spełnienia wymogi (Brzeziński, 2006).

Podejmowane problemy badawcze rzadko mają prostą strukturę jednozmienną (obejmującą jedną bądź kilka zmiennych objaśniających o niezależnych poziomach

Elżbieta Aranowska, Wydział Psychologii, Uniwersytet Warszawski, ul. Stawki 5/7, 00–183 Warszawa.

Jolanta Rytel, Instytut Psychologii, Uniwersytet Kardynała Stefana Wyszyńskiego, ul. Wóycickiego 1/3, budynek 14, 01–938 Warszawa, e-mail: j.rytel@uksw.edu.pl

Autorki składają serdeczne podziękowania Panu Profesorowi Andrzejowi Strzałeckiemu za życzliwość oraz wyrażenie zgody na wykorzystanie wyników badań i udostępnienie części rezultatów przeprowadzonych analiz.

wartości oraz jedną tylko mierzalną zmienną analizowaną, czyli objaśnianą). W takiej sytuacji badacz powinien rozważyć zastosowanie modelu jednowymiarowej, jedno- bądź wieloczynnikowej analizy wariancji (ANOVA), jako że populacje wyróżnione przez niezależne poziomy zmiennych objaśniających to rozłączne grupy obiektów stanowiące populacje niezależne. W przebiegu procesu badawczego w naukach empirycznych o wyborze modeli statystycznej analizy danych należy myśleć już na etapie operacjonalizowania analizowanych konstruktów, a brak rzetelnej wiedzy na temat bardziej zaawansowanych, wielowymiarowych modeli analizy może nie tylko utrudnić, ale wręcz uniemożliwić badaczowi właściwe – na poziomie treściowym – sformułowanie problemu.

W podstawowych podręcznikach do nauczania statystyki jej przydatność uzasadnia się koniecznością ustalenia optymalnego schematu badawczego do istniejących już hipotez. Trudno się z tym zgodzić! Gdyby przyjąć taki punkt widzenia, nowe metody analizy danych nie mogłyby ani powstawać, ani się rozwijać. Co gorsza, brak znajomości metod wielowymiarowych może ograniczać możliwość dostrzeżenia przez badacza wielowymiarowej struktury podejmowanego problemu i jej specyfiki. Innymi słowy, statystyka determinuje twórcze stawianie hipotez badawczych. Taki „wielowymiarowy sposób myślenia” już ponad ćwierć wieku temu postulowali Huberty i Smith (1982, s. 429).

Model wielowymiarowej, jedno- lub wieloczynnikowej analizy wariancji (MANOVA) stosuje się zwykle, analizując co najmniej dwie lub więcej mierzalnych zmiennych objaśnianych, przy czym analizowane populacje, identyfikowane przez poziomy zmiennych objaśniających, to wyłącznie populacje niezależne. W takiej sytuacji powstaje pytanie o wybór optymalnego modelu statystycznej analizy danych: MANOVA czy wielokrotne analizy jednowymiarowe (ANOVA)? Choć wybór wielowymiarowej analizy wariancji wydaje się oczywisty – badacz analizuje więcej niż jedną zmienną objaśnianą – to w specyficznych przypadkach lepszym wyborem byłoby przeprowadzenie wielu analiz jednowymiarowych. Dzieje się tak na przykład wtedy (Huberty i Morris, 1989):

- gdy zmienne objaśniane nie są ze sobą powiązane merytorycznie, toteż można traktować je jako pojęciowo niezależne (na przykład: wyobrażenia przestrzenna, zadowolenie z życia i perseweratywność). Oznacza to, że bardziej ogólna struktura latentna wobec każdej z tych zmiennych nie jest (ale może być) przedmiotem zainteresowania badacza;

- gdy badanie ma charakter eksploracyjny, a celem badacza jest sprawdzenie statusu metodologicznego nowej, wprowadzonej zmiennej objaśnianej; chodzi zatem

o efekty oddziaływań czynnika/czynników na tę właśnie zmienną;

- gdy wszystkie zmienne objaśniane lub ich części były uprzednio analizowane w kontekście jednowymiarowym – oczywiste jest, że badacz w badaniach porównawczych uwzględni ten sam kontekst;

- gdy celem badacza jest wyselekcjonowanie grup o podobnej charakterystyce ze względu na zmienne objaśniane, co stanowi wymóg konieczny do zaplanowania dalszych badań o charakterze eksperymentalnym.

Model MANOVA staje się optymalnym wyborem, gdy badacz traktuje mierzalne zmienne objaśniane jako powiązane ze sobą, a tym samym tworzące strukturę relacyjną cech, którą sensownie – w całości lub w części – da się interpretować merytorycznie. Implikuje to potrzebę wyłonienia (ewentualnie) ich podzbioru możliwie najsilniej różnicującego analizowane populacje niezależne (wskaźniki przez poziomy zmiennych objaśniających) oraz określenia wkładu każdej ze zmiennych w dyskryminację między populacjami. Wspomniane aspekty określić można, odpowiednio, jako problem selekcji zmiennych i problem ich uporządkowania (Huberty i Morris, 1989).

Równocześnie warto zauważyć, że dla decyzji o wyborze modelu analizy nie bez znaczenia jest to, jaki system reprezentują powiązane ze sobą zmienne. Badacz może traktować zmienne obserwowalne jako (powiązane ze sobą) przyczynowe wskaźniki pewnej zmiennej (lub zmiennych) latentnej, którą stara się ujawnić czy odkryć. Ale może także przyjąć, że są one skutkiem empirycznym, czyli sposobem, w jaki manifestuje się owa zmienna latentna. W tym ostatnim przypadku badacz winien rozważyć wykorzystanie modelu równań strukturalnych, które stanowią lepsze rozwiązanie niż model MANOVA (Bollen i Lennox, 1991; Cole, Maxwell, Arvey i Salas, 1993a).

We wstępie do pierwszego wydania *Wielowymiarowej analizy statystycznej* Donald F. Morrison (1990, s. 11) zauważa, że „w wielowymiarowej analizie statystycznej stykamy się z obserwacjami wielu zmiennych dokonywanymi na tym samym obiekcie. Takie właśnie obserwacje spotykamy często w socjologii, naukach społecznych, behawioralnych oraz w medycynie: ceny towarów, czasy reakcji zdrowego organizmu czy też wyniki badania krwi u pacjenta są przykładami danych wielowymiarowych. Tak jak w statystyce jednowymiarowej, zakładamy, że próba składa się z wektorów obserwacji dokonywanych na różnych osobach czy też innych niezależnych od siebie obiektach. Niemniej jednak wspólne pochodzenie zmiennych w jednym wektorze obserwacji powoduje, iż zmienne te są zależne. Jest to właśnie ta cecha, która odróżnia analizę wielowymiarową od jej jednowymiarowego pierwowzoru”.

Równocześnie nie bez znaczenia jest, jaki status metodologiczny mają zmienne obserwowane na tym samym obiekcie, w konsekwencji – na tej samej populacji: mogą być to częściowo lub całkowicie treściowo różne zmienne, jedna zmienna badana wielokrotnie w zmieniających się warunkach lub jedna zmienna badana wielokrotnie w tych samych warunkach. W przypadku dwu ostatnich sytuacji mówi się o powtarzalności pomiaru, ale w dwu różnych aspektach:

„– powtarzalność doświadczeń w tych samych warunkach (badanie tego samego w identyczny sposób), która na poziomie pomiarowym, czyli wyniku mierzenia, nazywana jest replikacją (*replication*) oraz

– powtarzalność doświadczeń w różnych warunkach (badanie tego samego przy założeniu działania różnych poziomów zmiennych niezależnych), którą na poziomie pomiarowym – wobec braku polskiego określenia – można by nazwać repetycją (*repeated measures*)” (Aranowska, 2005, s. 12). Ilustrację wspomnianych wyżej trzech sytuacji stanowi Rysunek 1.

Ze względu na postać hipotez zerowych właściwych dla całej klasy metod analizy wariancji, mianowicie równość wartości oczekiwanych (średnich) pewnych populacji, wyłącznie w sytuacji (b) możliwe jest traktowanie planu badawczego jako najprostszego schematu analizy wariancji z całkowicie powtarzanimi pomiarami dla k-zależnych populacji (inaczej jednoczynnikowej analizy wariancji z całkowicie powtarzanimi pomiarami).

W sytuacji (a) – z powodu różnic merytorycznych między analizowanymi zmiennymi lub też prawdopodobnej różnicy długości skal zmiennych – porównywanie wartości oczekiwanych jest bezzasadne; badacz powinien zainteresować się raczej metodami korelacyjnymi (na przykład takimi, jak modele regresji, eksploracyjne lub

konfirmacyjne analizy czynnikowe, analizy kanoniczne czy modele równań strukturalnych).

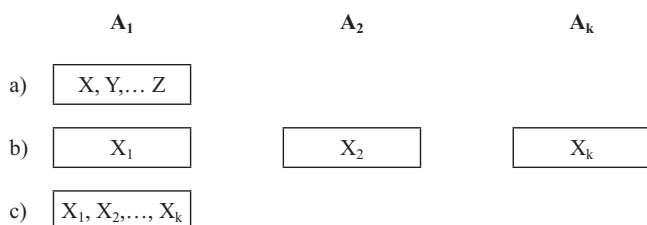
W sytuacji (c), mimo że długość skali przy każdym mierzeniu jest identyczna, porównywanie wartości oczekiwanych nie jest konieczne, ponieważ – z subiektywnego założenia badacza – nie zmieniły się warunki pomiaru, czyli kilka wartości pomiarowych – replikacji – uzyskuje się w obrębie tego samego poziomu czynnika. Ewentualne zmiany wartości pomiarowych u konkretnej osoby nie mogą być zatem tłumaczone przez działanie różnych poziomów czynnika kontrolowanego w badaniu, ale przez bliżej nieznaną fluktuację bodźców losowych, wprowadzające dyspersję wartości pomiarowych, którą można ocenić, obliczając na przykład ich wariancję, i traktować jako błąd losowy, czy – nazywając krócej – błąd, co nie tylko czyni się we wszystkich modelach analizy wariancji z powtarzanimi pomiarami, ale – co więcej – dodatkowo zaleca uwzględnianie w schematach badań replikacji pomiarów, aby móc ten błąd szacować.

Po wprowadzeniu co najmniej jednego jeszcze poziomu czynnika kontrolowanego w badaniu, poziomu zmiennej objaśniającej, plan badawczy w sytuacji (a) i (c) zmienia się na tyle, że możliwe będzie zastosowanie do analizy danych metod analizy wariancji. Jednak jaka to będzie metoda, zależy od tego, czy nowy poziom ukonstytuuje populacje zależne względem poprzednich, czy niezależne, jak pokazano na Rysunku 2.

Na Rysunku 2 w dwu kolumnach – sytuacja (a) i (c) – przedstawiono cztery schematy badawcze.

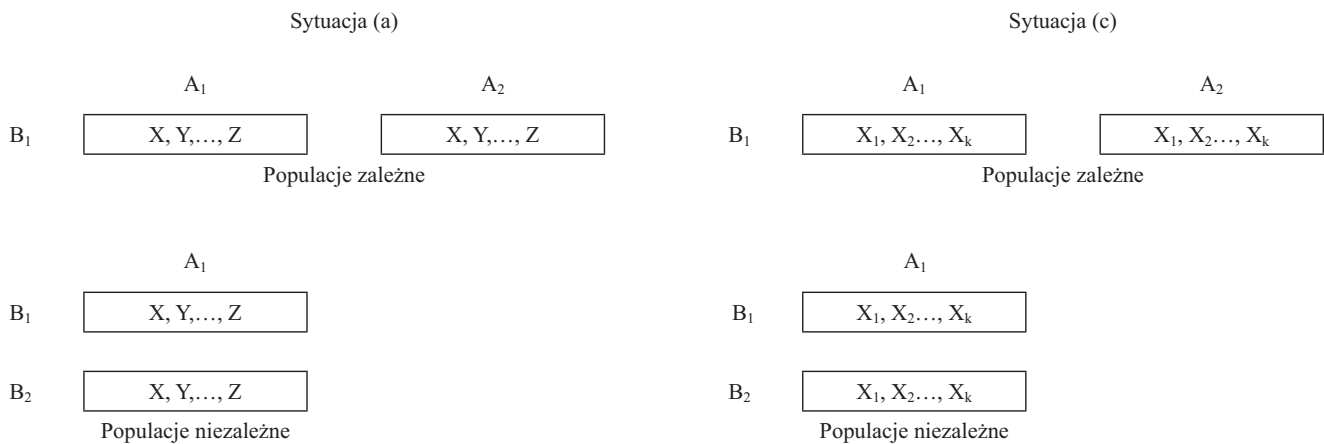
1) Tylko dolny rysunek sytuacji (a) stanowi podstawowy schemat dla MANOVA. Analiza danych z takich badań odbywałaby się poprzez zastosowanie k-wymiarowej (gdy różnych zmiennych mierzalnych: X, Y, ..., Z jest równa k) i jednoczynnikowej (czynnik B o dwu niezależnych poziomach wartości) analizy wariancji. Aby przeprowadzić obliczenia tą metodą, należy sprawdzić prawdziwość wszystkich jej założeń (opisanych niżej).

2) Na rysunku górnym w sytuacji (a) zaznaczono schematycznie dwukrotne badanie tej samej grupy osób wieloma różnymi zmiennymi mierzalnymi. Z jednej strony wiadomo, że porównywanie ich wartości przeciętnych jest albo bezzasadne (gdy długość skal jest inna), albo mało informacyjne (gdy skale są tej samej długości, ale treści – konstruktów – dość różne). Z drugiej zaś strony istnieją dwa powtarzane pomiary stanowiące populacje zależne dla każdej zmiennej oddzielnie. Analiza takich danych może przebiegać na wiele sposobów, w zależności od pytania badawczego. Podobnie jak w sytuacji (a) na Rysunku 1, stosowane tu metody statystyczne będą głównie należały do klasy metod stanowiących tzw. ogólnie



Rysunek 1.

Trzy podstawowe aspekty powtarzalności pomiaru: (a) skończona liczba różnych zmiennych objaśnianych mierzonych w tych samych ustalonych warunkach; (b) jedna konkretna zmienna objaśniana mierzona w różnych warunkach (repetycje po poziomach czynnika A); (c) jedna konkretna zmienna objaśniana mierzona skończoną liczbą razy w tych samych warunkach (replikacje).



Rysunek 2.

Rozwinięcie podstawowych planów badawczych dla powtarzanych pomiarów przedstawionych na Rysunku 1 w sytuacji (a) i (c) poprzez wprowadzenie dodatkowego poziomu czynnika odpowiadającego nowym warunkom badawczym.

ny model liniowy (GLM – por. np. Cohen, 1968; Graham, 2008). Tak więc np.:

- można analizować stabilność pomiaru poprzez określenie wartości i istotności korelacji r -Pearsona dla dwu pomiarów każdej zmiennej oddzielnie;

- można zbudować prosty model regresji objaśniający drugi pomiar każdej zmiennej oddzielnie przez pomiar pierwszy;

- można znaleźć różnicę, D , obydwu pomiarów dla każdej zmiennej oddzielnie i sprawdzać testem t -Studenta, czy średnia zmiennej D równa jest zeru (czy nie „przesunęły się” wartości oczekiwane – środki ciężkości skal – w drugim badaniu);

- można sprawdzać, czy wektor średnich zmiennych D dla wszystkich zmiennych analizowanych łącznie jest wektorem zerowym. Jest to sprawdzenie braku oddziaływania nowych warunków badania na wartości przeciętne całej struktury analizowanych zmiennych łącznie. Warto pamiętać, że dzięki ciągle niepoprawionemu błędowi w oprogramowaniu pakietu SPSS hipotezę taką weryfikuje się przez uruchomienie w ogólnym modelu liniowym opcji z powtarzaniem pomiarów po jednym czynniku dla najprostszego modelu mieszanego analizy wariancji i użycie testu efektów między obiektami (*tests of between-subjects effects*) (Aranowska, 2005);

- można sprawdzać wymiarowość (merytoryczną i formalną) wszystkich zmiennych X, Y, \dots, Z metodami eksploracyjnej lub confirmacyjnej analizy czynnikowej, czy stosując analizę kanoniczną (*canonical analysis*) (por. np. Harris, 1975);

- można sprawdzać sieci, relacje powiązań między zmiennymi X, Y, \dots, Z metodami analizy ścieżkowej

(*path analysis*), wplatając w ścieżki powtórzone pomiary, lub tworząc ścieżki ze zmiennych różnic pomiarów, D . Warto zauważyć, że pojęcie korelacji kanonicznej i analizy ścieżek uznać można za podstawy, na których wyrosła nowoczesna klasa metod statystycznych wspomnianych niżej, SEM;

- można sprawdzać istnienie cech ukrytych determinujących nasilenie cech obserwowalnych X, Y, \dots, Z , tworząc model ich wzajemnych relacji, wykorzystując modelowanie równań strukturalnych (*structural equation modeling* – SEM) (Graham, 2008) lub szczególnie intensywnie ostatnio rozwijaną podklasę tych modeli, czyli modelowanie latentnych krzywych rozwojowych (*latent growth curve modeling* – LGM) (Konarski, 2004; Preacher, Wichman, MacCallum i Briggs, 2008).

Nie wyczerpano zapewne wszystkich możliwości analizy danych uzyskanych po użyciu omawianego wyżej schematu badań dla sytuacji (a), bowiem mnóstwo specyficznych pytań badawczych może uruchomić taki plan gromadzenia danych. Podobnie rzecz się ma z planem omówionym wcześniej, gdzie podstawową metodą analizy wydaje się MANOVA. To zawsze od decyzji badacza zależy postać hipotez i najlepszy sposób ich weryfikacji.

Niżej omówione zostaną dwa schematy umieszczone w drugiej kolumnie Rysunku 2 nazwanej sytuacją (c).

3) Na rysunku górnym zmienna replikowana k -krotnie mierzona jest dwa razy.

Zgodnie z rozumieniem replikacji uzyskane wyniki stanowią ponowne, zwielokrotnione pomiary nasilenia cechy w tych samych warunkach dla uwiarygodnienia rzetelności operacji mierzenia. Wystarczy w tym przypadku jakąś miarę syntetyczną z pierwszej grupy replikacji po-

równać z tą miarą dla drugiej ich grupy, np. dwie wartości oczekiwane (testem *t*-Studenta dla populacji zależnych) czy dwie wariancje (testem Fergusona). Oczywiście, jeśli mimo operowania takim schematem badacz nie będzie chciał traktować swoich wyników jako grup replikacji odwzorowujących dokładność pomiarów (w dwu różnych warunkach) i będzie optował przy traktowaniu ich jako powtarzanych pomiarów, może zastosować do porównywania wartości oczekiwanych – ze szkodą dla użytego rozwiązania formalnego, a więc i dla interpretacji treściowej – dwuczynnikową analizę wariancji z całkowicie powtarzanymi pomiarami (po dwu poziomach czynnika A oraz po *k* poziomach rozłożonego w nim innego, nowego czynnika). Brak uwzględnienia replikacji w planach badań, których wyniki można analizować metodami powtarzanych pomiarów, uniemożliwia weryfikowanie niektórych spośród dopuszczalnych hipotez zerowych oraz powoduje, że statystyki *F*, stosowane do sprawdzenia części pozostałych, możliwych do zweryfikowania hipotez, charakteryzują się nieadekwatną – wobec liczników – postacią mianownika (por. np. Brzeziński i Stachowski, 1984).

4) Na dolnym schemacie w sytuacji (c) na Rysunku 2 wprowadzone niezależne dwa poziomy czynnika B implikują zbieranie replikowanych *k*-krotnie wyników jednej zmiennej w tych samych warunkach (jeden poziom czynnika A) w dwóch próbach niezależnych. Podobnie jak w punkcie wyżej, nieuwzględnianie zmienności między replikacjami może prowadzić – po wyciągnięciu z nich średnich – do analizy danych testem *t*-Studenta dla dwu populacji niezależnych. Natomiast ignorowanie narzuconego przez samego badacza schematu z replikacjami i uznawanie ich za powtarzane pomiary prowadzi go w tym przypadku do stosowania przy analizie danych np. dwuczynnikowej analizy wariancji z częściowo powtarzanymi pomiarami (po jakimś czynniku nowym, działającym na *k* poziomach zależnych, gdy czynnik B natomiast działa na dwóch poziomach niezależnych).

Dokładanie nowych poziomów zaznaczonych na Rysunku 2 czynników A i B zasadniczo nie zmienia klas proponowanych analiz, natomiast dokładanie nowych poziomów i nowych czynników (C, D, ...) zwiększy wymiarowość przestrzenną planów MANOVA czy powtarzanych pomiarów (a zatem liczbę dopuszczalnych hipotez, liczbę niezbędnych założeń, łączną liczbę prób i – czasami – liczbę osób badanych w pojedynczej pod-próbie).

MANOVA stosuje się najczęściej wtedy, gdy próbuje się sprawdzać położenie wartości oczekiwanych (średnich) grupy zmiennych mierzalnych łącznie w różnych populacjach niezależnych (np. w różnych grupach osób). Jeżeli

te położenia nie różnią się, nie ma podstaw do uznania, że przyczyna podziału populacji na podpopulacje (efekty oddziaływania jakiegoś czynnika) realnie determinuje przesunięcia średnich.

Analizę wariancji z powtarzanymi pomiarami (całkowicie bądź częściowo) stosuje się najczęściej wtedy, gdy próbuje się sprawdzać położenie wartości oczekiwanej (średniej) jednej zmiennej mierzalnej w zmieniających się warunkach badania, a więc głównie w różnych populacjach zależnych (np. w tej samej grupie osób). Dodanie w schematach badawczych – do przynajmniej jednego czynnika o poziomach zależnych – nowego czynnika o poziomach niezależnych zmienia charakter planów z pomiarów całkowicie powtarzanych na powtarzane częściowo.

Wymienione wyżej dwie duże klasy metod mają różne założenia formalne. Założenia metod dla powtarzanych pomiarów są mocniejsze. Główne różnice wynikają z różnego traktowania wyników pojedynczego elementu próby (osoby). W planie MANOVA *i*-ta osoba w konkretnych warunkach badania (przy konkretnym poziomie czynnika) charakteryzuje się wartościami: x_i, y_i, \dots, z_i . Liczby te określają nasilenia różnych zmiennych poprzez swoje położenie w pewnych miejscach na skalach tych zmiennych, skalach być może o różnej długości. Zatem wartość 8 na skali X wcale nie musi być większa od wartości 6 na skali Y, jeśli tylko skala Y jest dostatecznie krótsza (nasilenie cechy odpowiadające pomiarowi równemu 8 przy skali 40-punktowej jest znacznie mniejsze, niż nasilenie cechy odpowiadające pomiarowi równemu 6 przy skali 20-punktowej; wystarczy w tym celu porównać ułamek $8/40 = 0,2$ z ułamkiem $6/20 = 0,3$). Dlatego nie ma możliwości operowania pojęciem zmienności dla ciągu wartości x_1, y_1, \dots, z_1 . Konsekwentnie bezsensowne jest liczenie wartości jakiegokolwiek miary rozproszenia, np. wariancji tych pomiarów.

Odwrotnie w sytuacji pomiarów powtarzanych: obojętnie musimy oceniać zmienność wartości *i*-tej osoby: $x_{1i}, x_{2i}, \dots, x_{ki}$, by testować hipotezę o dynamice średniej zmiennej X w różnych, *k*, warunkach badania. Wariancja wyników pojedynczego elementu próby estymuje wariancję populacji tych wszystkich elementów, które reagują na zmieniające się warunki, tak jak ten element.

Uwzględnianie bądź nieuwzględnianie pewnych zmienności pomiarów w różnych metodach statystycznych odróżnia je od siebie między innymi na poziomie treści założeń i – w konsekwencji – przyjmowanych rozwiązań formalnych. Podsumowując, nie ma możliwości zastępowania jednej metody statystycznej inną bez straty jakichś informacji czy zysku innych. Naturalnie wpływa to na interpretację merytoryczną wyników tych metod.

Niezależność analizy wartości oczekiwanych i analizy zależności między zmiennymi mierzalnymi

Analiza struktury powiązań między zmiennymi objaśnianymi jest absolutną powinnością każdego badacza przynajmniej z dwóch ważnych powodów. Pierwszy z nich jest natury merytorycznej, drugi zaś – formalnej. W empirii rzadko ma się do czynienia ze zmiennymi, które są ze sobą parami nieskorelowane, dlatego badacz bezwzględnie musi określić siłę i kierunki powiązań tych zmiennych, ponieważ to właśnie zakres skorelowania między zmiennymi warunkuje możliwość analizowania i interpretowania różnic między wartościami oczekiwanymi tych zmiennych w różnych populacjach. Poniższy prosty przykład ilustrujący podstawowy schemat dla metody MANOVA (dwie zmienne objaśniane o różnym stopniu skorelowania oraz jedna zmienna objaśniająca o dwóch niezależnych poziomach wartości), dobrze dokumentuje tę tezę.

Przykład 1. W dwóch 5-osobowych, losowych, niezależnych próbach dokonano pomiaru dwóch zmiennych objaśnianych Y_1 i Y_2 , mierzonych na skalach o różnej długości. Rozpatrzmy – teoretycznie możliwe – trzy sytuacje. W każdej z nich inny jest stopień skorelowania zmiennych objaśnianych (współczynnik korelacji pomiędzy zmiennymi Y_1 i Y_2 dla każdej z trzech sytuacji oznaczony jest symbolem r z indeksem $i = 1, 2$, oznaczającym odpowiednio obie grupy, zaś symbolem \bar{r} oznaczono uśrednioną korelację dla obu grup), natomiast takie same pozostają średnie zmiennych objaśnianych Y_1 i Y_2 (oznaczone jako \bar{y}) oraz ich wariancje (oznaczone jako s^2) i odchylenia standardowe (oznaczone jako s).

Wprowadzenie do eksperymentu dwu zmiennych objaśnianych oznacza, że z pewnego punktu widzenia właśnie one obydwie opisują to zjawisko, które badacz chce szczegółowo analizować. W jakim aspekcie zatem badacz będzie myślał o wpływie (w sensie funkcjonalnym, a nie przyczynowym) zmiennej objaśniającej na badane zjawisko? Logika nasuwa odpowiedź: będzie on myślał o wpływie czynnika na całą strukturę zmiennych objaśnianych, czyli o łącznym wpływie na obydwie zmienne, a nie na każdą z nich oddzielnie. Jeżeli przy przechodzeniu z jednego poziomu czynnika na drugi, średnie zmiennych objaśnianych nie tylko nie pozostają w tych samych miejscach na skali zmiennej, ale zmieniają się w tych samych kierunkach we wszystkich zmiennych, gdy zmienne są nieskorelowane, to wpływ czynnika jest znacznie większy na tę strukturę niż wtedy, gdy tylko część zmiennych zachowuje się według tej zasady lub wtedy, gdy są one skorelowane. Jeśli nie ma wspólnej dla całej grupy reguły przyporządkowań między wartościami zmiennych

objaśnianych i przyporządkowania te są przypadkowe (czyli wartość współczynnika korelacji $\rho = 0$) tak w całej populacji, jak i w podpopulacjach wyznaczonych przez poziomy zmiennej objaśniającej, to wpływ czynnika jest znaczny, jeśli jego efektem jest przesunięcie wartości wszystkich zmiennych w tym samym kierunku z poziomu na poziom. Toteż wartość testu weryfikującego hipotezę zerową o braku wpływu zmiennej objaśniającej na grupę zmiennych objaśnianych, przy tej samej różnicy średnich w próbach, powinna być wyższa przy braku korelacji między zmiennymi niż przy jej istnieniu (por. Tabachnik i Fidell, 2001).

Z tego powodu różnice między średnimi grupowymi we wszystkich trzech przedstawionych powyżej sytuacjach, mimo iż są identyczne dla zmiennej Y_1 i identyczne dla zmiennej Y_2 , powinny być traktowane jako coraz słabsze przy coraz większym skorelowaniu tych zmiennych. Zatem wartość testu weryfikującego hipotezę o braku wpływu czynnika na strukturę zmiennych powinna być najwyższa w sytuacji I oraz odpowiednio niższa w pozostałych sytuacjach.

Rozważmy sytuację, w której korelacje między zmiennymi w populacji i wszystkich podpopulacjach są maksymalne. To, że zmienne są od siebie liniowo zależne, nie ma wpływu na stopień ani kierunek zmian średnich przy zmianie poziomów czynnika: (a) różnica może być nieistotna statystycznie, (b) może być istotna, a średnie dla obydwu zmiennych zgodnie maleją bądź rosną, (c) może być istotna, jednakże średnie dla jednej zmiennej wzrosły, a dla drugiej zmaleły. Jednorodny wpływ czynnika wystąpił tylko w sytuacji (b) i ją właśnie chcielibyśmy premiować, wyrażając to większymi wartościami testu weryfikującego hipotezę o braku wpływu czynnika. Toteż wartość testu weryfikującego hipotezę zerową o braku wpływu czynnika na grupę zmiennych, przy tej samej i w tych samych kierunkach różnicy średnich w próbach, powinna być wyższa przy bardzo silnej korelacji między zmiennymi niż wtedy, gdy tylko jedna zmienna częściowo wyjaśnia drugą.

Z powyższych rozważań wynika niezwykle ważny wniosek. Otóż nasze przewidywania o zmianach tendencji centralnych zmiennych mają sens wyłącznie wtedy, gdy założymy stałą strukturę powiązań między tymi zmiennymi w populacji i podpopulacjach wyznaczonych przez poziomy czynników. Gdyby zupełnie dowolnie zmieniały się powiązania zmiennych przy przechodzeniu z poziomu na poziom czynników, nasze spekulacje dotyczące średnich nie miałyby żadnego sensu oraz praktycznie niemożliwe stałoby się skonstruowanie testu ujawniającego te wszystkie zmiany.

Tabela 1.
 Niezależność współczynnika siły związku dwu zmiennych od miar tendencji centralnej

Grupy	Lp.	Sytuacja 1		Sytuacja 2		Sytuacja 3	
		Y ₁	Y ₂	Y ₁	Y ₂	Y ₁	Y ₂
Grupa I	1	6	100	6	100	6	100
	2	5	70	5	90	5	90
	3	6	40	6	100	6	100
	4	4	90	4	70	4	70
	5	4	100	4	40	4	40
	Σ	25	400	25	400	25	400
	\bar{y}_1	5	80	5	80	5	80
	s_1^2	$\frac{4}{4}$	$\frac{2600}{4}$	$\frac{4}{4}$	$\frac{2600}{4}$	$\frac{4}{4}$	$\frac{2600}{4}$
	s_1	1	25,5	1	25,5	1	25,5
	r_1	-0,4903	$\left(\text{cov} = \frac{-50}{4}\right)$	0,8825	$\left(\text{cov} = \frac{90}{4}\right)$	0,8825	$\left(\text{cov} = \frac{90}{4}\right)$
Grupa II	6	8	110	8	80	8	110
	7	7	100	7	110	7	100
	8	8	80	8	110	8	110
	9	6	100	6	100	6	80
	10	6	110	6	100	6	100
	Σ	35	500	35	500	35	500
	\bar{y}_2	7	100	7	100	7	100
	s_2^2	$\frac{4}{4}$	$\frac{600}{4}$	$\frac{4}{4}$	$\frac{600}{4}$	$\frac{4}{4}$	$\frac{600}{4}$
	s_2	1	12,25	1	12,25	1	12,25
	r_2	-0,4083	$\text{cov} = \frac{-20}{4}$	-0,2041	$\text{cov} = \frac{-10}{4}$	0,8165	$\text{cov} = \frac{40}{4}$
Estymatory nieobciążone (z dwu prób n_1 i n_2)	\bar{y}	6	90	6	90	6	90
	\bar{s}	1	20	1	20	1	20
	\bar{r}	-0,4375		0,5000		0,8125	

Toteż oprócz równości wariancji każdej zmiennej objaśnianej we wszystkich podpopulacjach – to założenie znamy z analizy wariancji jednowymiarowej (ANOVA) – podstawowym, drugim i równie ważnym założeniem wielowymiarowej analizy wariancji jest identyczność korelacji albo – co jest warunkiem równoważnym – identyczność kowariancji każdej pary zmiennych objaśnianych we wszystkich podpopulacjach wyznaczonych przez poziomy czynników kontrolowanych w badaniu.

Jaką wartość (jeszcze nieznaną) statystyki weryfikującej brak efektów oddziaływania czynnika na zmienne objaśniane spodziewamy się uzyskać w każdej z trzech rozważanych powyżej sytuacji (por. Tabela 1)? Na samym dole Tabeli 1 są podane odchylenia standardowe zmiennej Y_1 i Y_2 . Otóż średnie zmiennej Y_1 różnią się o dwa odchylenia standardowe populacji ($\sigma_1 \approx \bar{s}_1 = 1$), przy czym średnia w drugiej grupie jest wyższa. Podobnie, średnie zmiennej Y_2 różnią się dokładnie o jedno odchylenie

standardowe populacji ($\sigma_2 \approx \bar{s}_2 = 20$), średnia w grupie drugiej jest również wyższa. Reasumując, średnie obu zmiennych Y_1 i Y_2 różnią się w grupie pierwszej i drugiej – zważywszy na identyczną liczbę badanych w każdej próbie – prawie tak samo w sensie liczby błędów standardowych – oraz $\bar{x}_2 > \bar{x}_1$ dla obu zmiennych. Uwzględniając zatem związek tych zmiennych, spodziewamy się wyższych wartości statystyki weryfikującej hipotezę o braku wpływu czynnika w sytuacji I i III oraz niższej wartości tej statystyki w sytuacji II (rozwiązanie tego przykładu przedstawione zostanie dalej).

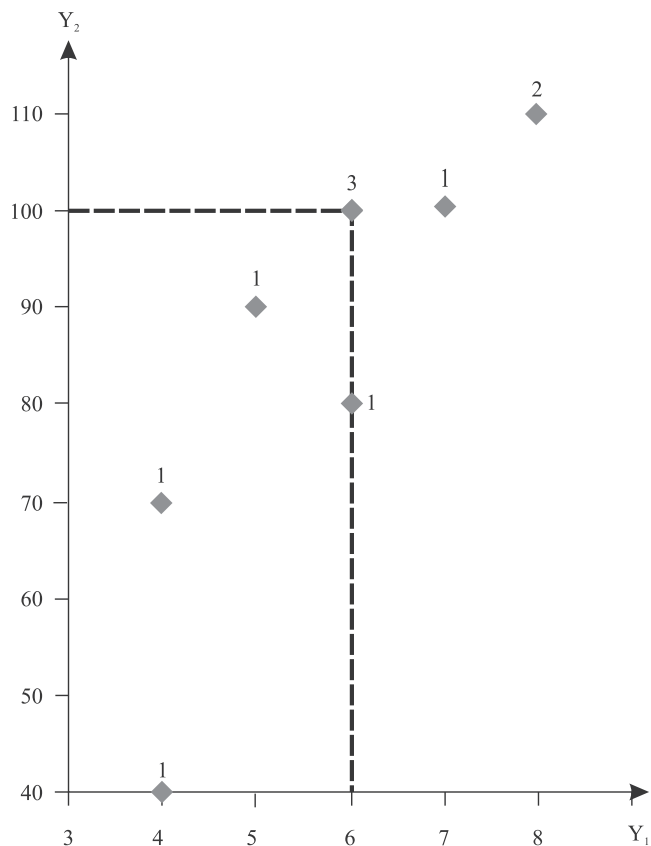
Założenia analizy wielowymiarowej

Każdej osobie badanej przyporządkowanych jest p wartości pomiarowych, które odpowiadają określonym zmiennym objaśnianym: Y_1, Y_2, \dots, Y_p (w przykładzie 1 $p = 2$). Dla każdego badanego otrzymaliśmy zatem p -wymiarowy wektor obserwacji¹. Gdybyśmy potrafili wyobrazić sobie układ współrzędnych składający się z p osi, to ten zbiór p liczb dla jednej osoby określałby dokładnie jeden punkt p -wymiarowej przestrzeni, równoznaczny z końcem wektora obserwacji (początkiem wektora jest początek układu).

W przykładzie 1 jest to przestrzeń dwuwymiarowa i łącznie wszystkie te punkty – końce wektorów – odtworzyć dla dziesięciu osób. Zrobimy to dla sytuacji III z przykładu 1. Na Rysunku 3 widać, że dla każdej osoby nie są to punkty różne, np. punkt (6, 100) wystąpił aż trzy razy, co zostało zaznaczone liczbami znajdującymi się obok każdego z nich. Zamiast tych liczb, można byłoby wprowadzić trzecią oś, oś frekwencji i punktowi (6, 100) przyporządkować trzy jednostki z tej osi. Powstałaby zatem już nie dwu-, ale trójwymiarowa przestrzeń (przy p zmiennych $p + 1$ -wymiarowa), w której ostatni wymiar zarezerwowany zostałby dla zliczania, ile razy dany punkt (a tym samym – wektor) wystąpił w badaniu.

Nie wiemy, jak wyglądałby taki rozkład frekwencji w populacji, nie wiemy też, które wartości występują najczęściej, a które najrzadziej. Toteż zamiast o frekwencjach przyporządkowanych dyskretnym parom liczb naturalnych odpowiadającym wartościom skal obydwu zmiennych – myśląc o populacji – mówimy (po niezbędnej idealizacji w postaci uciąglenia tych zmiennych) o prawdopodobieństwach, a precyzyjniej o funkcji gęstości prawdopodobieństwa dla dwu zmiennych łącznie, czyli o dwuzmiennowej funkcji gęstości prawdopodobieństwa.

Wielowymiarową analizą wariancji (MANOVA) można posługiwać się wtedy, gdy te p -wymiarowe wektory obserwacji pochodzą z populacji o p -wymiarowym rozkładzie normalnym prawdopodobieństwa. Założenie to jest zgodne z podobnym założeniem dla jednowymiaro-



Rysunek 3.

Geometryczny obraz wartości uzyskanych dla dwu zmiennych w próbie dziesięcioosobowej (przykład 1, sytuacja III).

wej analizy wariancji (ANOVA) dotyczącym normalności rozkładu prawdopodobieństwa zmiennej objaśnianej w populacji. Każdy taki p -wymiarowy wektor w przestrzeni cech ma więc określoną – przez rozkład normalny – szansę pojawienia się w naszej próbie, jeśli pochodzi ona z populacji o tym rozkładzie.

Zastanówmy się, co charakteryzuje ten rozkład. Rozkład jednej zmiennej jest scharakteryzowany przez średnią populacji oraz wariancję. Podobnie dla rozkładu zmiennej złożonej z wielu zmiennych – z p zmiennych, jak krótko powiemy dla (wektorowej) zmiennej p -wymiarowej – określone są dwa parametry charakteryzujące ten rozkład. Zamiast jednej średniej mamy teraz p średnich w populacji, jedną dla każdej zmiennej. Zatem pierwszym parametrem jego rozkładu jest wektor μ , równy:

$$(1) \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Drugi parametr rozkładu zależy nie tylko od wariancji wszystkich p zmiennych, ale także od stopnia powiązań między nimi. Jest nim macierz kowariancji Σ , postaci:

$$(2) \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

gdzie σ_{ii} oznacza wariancję w populacji każdej i -tej zmiennej Y_i , dla $i = 1, 2, \dots, p$, natomiast σ_{ij} oznacza kowariancję zmiennych Y_i z Y_j w populacji dla $i = 1, 2, \dots, p$, $j = 1, 2, \dots, p$ oraz $i \neq j$.

Przez analogię do rozkładu jednej zmiennej, drugi parametr rozkładu powinien być zależny od wariancji wszystkich badanych p zmiennych. Istotnie, takie wariancje znajdują się na głównej przekątnej macierzy (2)². W jaki sposób estymuje się postać macierzową tego parametru? Odpowiedź jest prosta: rozproszenie wyników w takiej p -wymiarowej przestrzeni zależy od stopnia powiązań zmiennych objaśnianych, czyli osi układu współrzędnych, na których „rozpina się” przestrzeń naszych wektorów obserwacji. Wystarczy przeanalizować ten fakt, odwołując się do sytuacji III omawianego przykładu 1. Średnia korelacja pomiędzy Y_1 i Y_2 (oznaczona jako \bar{r}) równa jest tam 0,8125 i jest istotna statystycznie. Można przypuszczać, że w populacji istnieje silny związek dodatni między Y_1 i Y_2 , co oznacza, że dużym wartościom zmiennej Y_1 odpowiadają – z reguły – duże wartości zmiennej Y_2 , małym wartościom Y_1 – małe wartości Y_2 . Szansa, że takie właśnie pary pojawią się w próbie, jest znacznie wyższa niż szansa, że pojawią się inne mieszane pary wartości, tzn. małe z dużymi. Zatem w rozważanej przestrzeni trójwymiarowej, w której dwie osie stanowią odpowiednio zmienne objaśniane Y_1 i Y_2 , a trzecią – prawdopodobieństwo realizacji wartości obydwu zmiennych łącznie, kształt funkcji gęstości dwuwymiarowego rozkładu normalnego zostanie wydłużony wzdłuż linii regresji między Y_1 i Y_2 (por. Koronacki i Mielniczuk, 2001, rys. 2.9, s. 133). Na tej linii bowiem są najwyższe prawdopodobieństwa dla par wartości Y_1 i Y_2 w najbliższym położeniu punktu (μ_1, μ_2). Im dalej od tego punktu na linii regresji, tym bardziej maleją prawdopodobieństwa zrealizowania par wartości obydwu zmiennych, ale i tak są one znacznie wyższe niż prawdopodobieństwa realizacji punktów nie należących na tej prostej.

Gdyby ściąć czubek takiego wydłużonego (przy bardzo silnej korelacji między zmiennymi) dwuwymiarowego normalnego rozkładu prawdopodobieństwa wzdłuż płaszczyzny równoległej do płaszczyzny Y_1, Y_2 , wtedy brze-

gi wyznaczonej warstwy byłyby bardzo wąską elipsą. Elipsa ta to linia tych samych wartości funkcji gęstości dwuwymiarowego rozkładu prawdopodobieństwa. Im bardziej maleje zależność między zmiennymi w populacji, tym bardziej te wąskie elipsy – odpowiadające tym samym wartościom funkcji gęstości prawdopodobieństwa – rozszerzają się, przyjmując postać okręgu w skrajnym przypadku, gdy korelacja równa jest dokładnie zeru.

Widać więc, że o rozrzutach wartości obu zmiennych nie decydują wyłącznie wariancje tych zmiennych, ale także siła powiązań między ich wartościami. Stąd dla p zmiennych drugi parametr rozkładu prawdopodobieństwa jest zależny i od wariancji, i od kowariancji par zmiennych.

Oczywiście nie znamy ani μ , ani Σ . Jednakże na podstawie próby możemy ustalić nieobciążone oceny obydwu parametrów populacji. Estymatorem nieobciążonym wektora μ jest wektor średnich w próbie (dla zmiennych Y_1, \dots, Y_p). W naszym przykładzie jest to wektor $\bar{y} = \begin{bmatrix} 6 \\ 90 \end{bmatrix}$.

Natomiast estymatorem nieobciążonym macierzy Σ jest macierz uśrednionych kowariancji S , postaci:

$$(3) \quad S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

gdzie na przekątnej znajdują się uśrednione wariancje zmiennych Y_1, \dots, Y_p z dwu badanych grup, natomiast symetrycznie powyżej i poniżej głównej przekątnej są uśrednione kowariancje par zmiennych, np. s_{1p} oznacza kowariancję zmiennej Y_1 i Y_p itd.³ W sytuacji III omawianego przykładu macierz S ma postać (por. wartości w Tabeli 1):

$$(4) \quad S = \begin{bmatrix} \frac{4+4}{8} & \frac{90+40}{8} \\ \frac{90+40}{8} & \frac{2600+600}{8} \end{bmatrix} = \begin{bmatrix} \frac{8}{8} & \frac{130}{8} \\ \frac{130}{8} & \frac{3200}{8} \end{bmatrix} = \begin{bmatrix} 1 & 16,25 \\ 16,25 & 400 \end{bmatrix}$$

Konkluzją jest główne założenie wielowymiarowej analizy wariancji: p -wymiarowe wektory obserwacji pochodzą z populacji, w której p -wymiarowa (wektorowa) zmienna losowa ma rozkład normalny z parametra-

mi μ , Σ , co oznaczamy krótko jako $N(\mu, \Sigma)$. Zmienna ta ma ponadto rozkład normalny w każdej z podpopulacji, wyznaczonej przez poziomy czynników i ich przecięcia, z parametrami np. dla i -tej podpopulacji równymi μ_i , Σ , czyli $N(\mu_i, \Sigma)$.

Założenie to zawiera wszystkie założenia jednowymiarowej analizy wariancji oraz jedno dodatkowe: założenie o identyczności kowariancji dla ustalonej pary zmiennych we wszystkich rozważanych podpopulacjach. Identyczność wariancji i kowariancji wyraża się tu w tej samej macierzy Σ .

Podsumujmy w punktach podane wyżej założenia MANOVA:

1. Wszystkie zmienne objaśniane łącznie muszą mieć wielowymiarowy rozkład normalny w populacji i wszystkich niezależnych podpopulacjach wyznaczonych przez poziomy czynników i ich przecięcia.

2. Każda zmienna objaśniana ma homogeniczne wariancje we wszystkich niezależnych podpopulacjach wyznaczonych przez poziomy czynników i ich przecięcia.

3. Konkretnie pary zmiennych objaśnianych mają kowariancje identyczne we wszystkich niezależnych podpopulacjach wyznaczonych przez poziomy czynników i ich przecięcia. Różne pary mogą oczywiście mieć różne kowariancje w danej podpopulacji. Założenie różności kowariancji dla różnych par zmiennych nie jest dopuszczalne np. w metodach z powtarzaniem pomiarami; jest to jedna z ważniejszych różnic między metodami.

4. Losowych, niezależnych (zewnątrznie i wewnątrznie) podrób jest tyle, ile wynosi iloczyn poziomów czynników kontrolowanych w badaniu (zmiennych objaśniających).

5. Wszystkie te podpróby są konieczne tej samej liczebności (w przeciwnym razie statystyka, którą intuicyjnie konstruujemy – przez analogię do przypadku jednowymiarowego – nie będzie miała żadnego ze znanych rozkładów prawdopodobieństwa, w szczególności tego, na którym nam zależy, czyli rozkładu prawdopodobieństwa F Fishera–Snedecora).

Sprawdzenie założeń przedstawionych wyżej jest konieczne. Prawdziwość założenia pierwszego nie do końca może być wykazana z powodu braku adekwatnego oprogramowania komputerowego. Dlatego badacze często w tej sytuacji odwołują się do teorii odporności testów statystycznych na pogwałcenie tego założenia (*robust theory*), która w przypadku jednowymiarowym orzeka, że wystarczy rozkład w miarę symetryczny i jednomodalny, niekoniecznie normalny, przy czym wielkość kurtozy, czyli stopnia spłaszczenia/smukłości rozkładu, nie może być znaczna.

W przypadku wielu zmiennych objaśnianych w praktyce badawczej próbuje się ominąć problem, analizując

choćby dwuzmiennowe rozkłady frekwencji próbkowych (Stevens, 2002, s. 263), naturalnie po wcześniejszym zadbaniu o znaczne rozmiary prób. Równocześnie rekomendowane jako praktyczne – bo łatwe do przeprowadzenia – rozwiązanie, polegające na sprawdzaniu normalności rozkładu każdej ze zmiennych objaśnianych z osobna (Field, 2009, s. 593), wydaje się mniej zasadne niż sprawdzanie normalności dwuwymiarowej dla każdej pary zmiennych. Normalność rozkładów dwuwymiarowych dostarcza informacji o najprostszym przypadku złożenia zmiennych partycypujących w całej wielowymiarowej strukturze, umożliwiając pierwszy wgląd w zachowanie zmiennych (czego nie da się orzec na podstawie stwierdzenia normalności rozkładów jednowymiarowych).

Gdy p -wymiarowy rozkład prawdopodobieństwa jest normalny, normalne są także rozkłady w przestrzeni o mniejszej liczbie wymiarów, ale nie odwrotnie. Jeżeli dwie zmienne, każda oddzielnie, mają normalny rozkład prawdopodobieństwa, to ich złożenie dwuzmiennowe nie tylko nie musi mieć rozkładu normalnego, ale gęstość dwuwymiarowa może nawet nie istnieć; innymi słowy z faktu, że istnieją rozkłady jednowymiarowe każdej z dwu zmiennych oddzielnie nie wynika, że będzie istniał ich rozkład łączny. Częściowo zależy to także od tego, czy zmienne są niezależne, czy zależne (Feller, 1978). W praktyce, gdy rozkłady brzegowe zmiennych objaśnianych można uznać za normalne, wielowymiarowy rozkład tych zmiennych często również jest rozkładem normalnym (Gnanadesikan, 1977, s. 168), a patologiczne przypadki danych, których reprezentacje dla mniejszej liczby wymiarów zachowują normalność, tracąc ją dla większej liczby wymiarów, w praktyce zdarzają się rzadko (Johnson i Wichern, 2002, s. 153). Pewne rozwiązanie stanowią testy dla skośności i kurtozy rozkładu wielowymiarowego (Mardia, 1980) bądź też proponowane w literaturze nowe metody, takie jak test sektorowy (*sector test of multinormality*), który stanowi uogólnienie na przypadek wielowymiarowy testu zgodności χ^2 (von Eye i Bogat, 2004, por. też Doornik i Hansen, 2008).

Założenia drugie i trzecie łącznie sprawdza się testem Boxa (por. np. Aranowska, 1996; Brzeziński i Stachowski, 1984) przeznaczonym do badania równości macierzy kowariancji w skończonej liczbie populacji niezależnych. Statystyka M tego testu stanowi uogólnienie testu Bartletta równości wariancji (Morrison, 1990). Już w 1953 roku Box wykazał, że test Bartletta jest bardzo wrażliwy na odstępstwa od normalności rozkładu. Należy zatem oczekiwać, że rozkład statystyki M także będzie bardzo czuły na pogwałcenie założenia o wielowymiarowej normalności zmiennych objaśnianych w analizowanych populacjach, na przykład prawdopodobieństwo przekroczenia przez

statystykę testu nominalnej pięcioprocentowej wartości krytycznej dla dwuwymiarowych rozkładów o małej kurtozie jest bardzo duże. Jak konkluduje Morrison (1990, s. 373): „omawiany test nie powinien być stosowany jako rozstrzygający dla danych, dla których wielowymiarowy rozkład normalny wydaje się nierealistyczny”. Założenie o równości macierzy kowariancji może zostać odrzucone za pomocą testu Boxa nie z powodu jego niespełnienia, ale z powodu odstępstw od normalności rozkładu wielowymiarowego zmiennych objaśnianych. Jako alternatywę badacz może wykorzystać techniki modelowania strukturalnego, umożliwiające sprawdzenie tego założenia za pomocą odpornych metod (Raykov, 2001). Może także zdecydować się przyjąć bardziej konserwatywny poziom istotności, np. $\alpha = 0,01$ (Raykov i Marcoulides, 2008), a spotyka się w literaturze rekomendacje, żeby przyjmować nawet $\alpha = 0,001$.

Prawdziwość założenia czwartego i piątego zależy wyłącznie od decyzji badacza. Jeśli nie zadbał w fazie planowania i prowadzenia badań o równoliczne próby, nie ma żadnych podstaw do analizy danych metodami MANOVA. Należy podkreślić, że założenie o niezależności obserwacji jest niezwykle ważnym, a często pomijanym w literaturze założeniem, na co zwraca uwagę Stevens (2002, s. 258). Nawet niewielka zależność pomiędzy obserwacjami powoduje dramatyczny, kilkakrotny wzrost rzeczywistego poziomu istotności w porównaniu z nominalnym, subiektywnie przyjętym przez badacza. W celu sprawdzenia tego założenia dla każdej zmiennej objaśnianej można wykorzystać współczynnik korelacji wewnątrzklasowej.

Postać hipotezy

Przez analogię do hipotezy jednowymiarowej, hipoteza wielowymiarowa dla omawianego przykładu będzie zakładała równość średnich, tutaj: dwóch zmiennych objaśnianych w dwóch podpopulacjach (wyróżnionych przez dwa niezależne poziomy czynnika) i w całej populacji, nadto będzie zapisana w postaci wektorowej:

$$(5) \quad H_0 : \mu_1 = \mu_2 = \mu$$

albo inaczej

$$(6) \quad H_0 : \alpha_i = 0 \text{ dla każdego } i \text{ (} i = 1, 2 \text{)}$$

gdzie α_i (w postaci wektorowej) standardowo oznacza efekt główny działania i -tego poziomu czynnika A na zmienną pierwszą i – odpowiednio – drugą.

Hipoteza zerowa dotyczy równości trzech wektorów w postaci:

$$(7) \quad \begin{bmatrix} \mu_{y_1,1} \\ \mu_{y_2,1} \end{bmatrix} = \begin{bmatrix} \mu_{y_1,2} \\ \mu_{y_2,2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

a więc równości odpowiadających sobie współrzędnych tych wektorów, czyli $\mu_{y_1,1} = \mu_{y_1,2} = \mu_1$ i równocześnie $\mu_{y_2,1} = \mu_{y_2,2} = \mu_2$. Zatem hipoteza zerowa dotyczy równości średnich zmiennej Y_1 we wszystkich podpopulacjach i całej populacji oraz jednocześnie równości średnich zmiennej Y_2 we wszystkich podpopulacjach i całej populacji. Uwzględniając definicję efektu głównego: $\mu_{y_1,1} - \mu_1 = \alpha_{y_1,1}$ oraz $\mu_{y_1,2} - \mu_1 = \alpha_{y_1,2}$ dla zmiennej Y_1 i podobnie dla Y_2 , hipotezę zerową można przedstawić jako:

$$(8) \quad \begin{bmatrix} \alpha_{y_1,1} \\ \alpha_{y_2,1} \end{bmatrix} = \begin{bmatrix} \alpha_{y_1,2} \\ \alpha_{y_2,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

lub też: $\alpha_{y_1,1} = \alpha_{y_1,2} = 0$ i równocześnie $\alpha_{y_2,1} = \alpha_{y_2,2} = 0$.

Widać, jak dalece zapis wektorowy uogólnia to, co chce się powiedzieć o wszystkich zmiennych badanych razem. Gdyby badacz objął analizą p ($p > 2$) zmiennych objaśnianych, przedstawione wyżej wektory składałyby się nie z dwu, lecz z p współrzędnych (wierszy).

Przy dwu czynnikach kontrolowanych w badaniu – podobnie jak w przypadku analizy jednowymiarowej – należy poddać sprawdzeniu trzy hipotezy z wektorami α_i , β_j oraz $(\alpha\beta)_{ij}$ równymi zeru. Dla wielu, c czynników, liczba weryfikowanych hipotez wynosi 2^c .

Postać statystyki

Zarówno w założeniach, jak i w postaciach hipotez zerowych występują analogie z analizą jednowymiarową. Fakt ten musi znaleźć odbicie w podobieństwie postaci statystyki. W przypadku wielowymiarowym mianownik statystyk stanowią uśrednione wariancje z podprób, tworzone równocześnie dla p zmiennych według zasady obowiązującej dla przypadku jednowymiarowego. Wiadomo, że w ANOVA mianownik statystyki F stanowi estymator łączny wariancji (zmiennej objaśnianej) w całej populacji utworzony przez uśrednienie wariancji próbkowych. Pojedyncza próba pobrana została losowo z podpopulacji wyznaczonej przez jeden poziom czynnika kontrolowanego w badaniu bądź – gdy jest ich więcej – przez przecięcie poziomów czynników. Przy wielu zmiennych objaśnianych rozrzut wyników w każdej najdrobniejszej podpróbce (wybranej z podpopulacji utworzonej przez przecięcia poziomów czynników) jest, jak wiadomo, zależny nie tylko od wariancji każdej ze zmiennych Y_1, \dots, Y_p oddzielnie, ale także od kowariancji między tymi zmiennymi. Tak więc dla każdej podpróby mamy jedną macierz, składającą się z wariancji i kowariancji między zmiennymi Y_1, \dots, Y_p

(wyliczonymi z pomiarów zawartych w tej podpróbie). Macierz wariancji-kowariancji łączną z wielu podprób tworzy się przez uśrednienie tych wszystkich macierzy (co pokazano wyżej). Na przykład przy kontrolowaniu dwu czynników A i B powstanie macierz – na wzór przypadku jednowymiarowego – z z $[k \cdot w \cdot (m - 1)]$ stopniami swobody, gdzie k to liczba poziomów czynnika A, w – liczba poziomów czynnika B, m – liczba osób badanych w każdej podpróbce, identyczna dla każdej z nich.

W przypadku jednowymiarowym była to średnia wariancja, teraz jest to średnia macierz. Macierz ta nosi oznaczenie S i jest interpretowana jako wielowymiarowy wynik ważenia wariancji p zmiennych niewyjaśnionych działaniem wprowadzonych do badania czynników. Macierz S jeszcze niepodzielona przez stopnie swobody jest oznaczana przez G . Dla przykładu 1 macierz

$$(9) \quad G = (n - k) \cdot S = \begin{bmatrix} 8 & 130 \\ 130 & 3200 \end{bmatrix}$$

gdzie $n = n_1 + n_2 = 5 + 5 = 10$ jest liczbą wszystkich badanych we wszystkich podpróbach łącznie, natomiast $k = 2$ i jest liczbą poziomów czynnika kontrolowanego w badaniu.

Podobnie jest z licznikiem poszukiwanej statystyki wielowymiarowej. Jego postać w analizie jednowymiarowej zależy od tego, która hipoteza zerowa jest weryfikowana. Licznik jest także wariancją i ma interpretację wariancji wyjaśnionej ewentualnym działaniem czynnika lub działaniem interakcji czynników. Na przykład w wypadku analizy dwuczynnikowej wariancja ta jest wyjaśniona albo przez wpływ czynnika A na Y , wtedy wariancja wyznaczona jest przez rozrzut średnich z poziomów tego czynnika wokół średniej globalnej (z całej próby), albo przez wpływ czynnika B na Y – wtedy wariancja wyznaczona jest przez rozrzut średnich z poziomów tego czynnika wokół średniej globalnej, albo przez wpływ interakcji czynników A i B na Y i wtedy wariancja w liczniku wyznaczona jest przez rozrzut interakcji.

Konstruując identyczne rozrzuty średnich z poziomów czynników czy ich interakcji przy problemie p zmiennych, zamiast jednej liczby (wariancji wyjaśnionej) otrzymuje się macierz, której elementy zależą wyłącznie od tych rozrzutów. Macierze te jeszcze niepodzielone przez stopnie swobody oznaczone są literą H z odpowiednimi indeksami, a więc: H_A , H_B , $H_{A \times B}$. Wypiszmy macierz H_A dla sytuacji III przykładu 1. Przez analogię z przypadkiem jednowymiarowym

$$(10) \quad H_A = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})'$$

gdzie \bar{x}_i oraz x są wektorami zawierającymi średnie z p cech, w naszym przykładzie dwóch cech.

Wybierając odpowiednie średnie z Tabeli 1, otrzymujemy

$$(11) \quad H_A = 5 \left(\begin{bmatrix} 5 \\ 80 \end{bmatrix} - \begin{bmatrix} 6 \\ 90 \end{bmatrix} \right) \left(\begin{bmatrix} 5 \\ 80 \end{bmatrix} - \begin{bmatrix} 6 \\ 90 \end{bmatrix} \right)' + \\ + 5 \left(\begin{bmatrix} 7 \\ 100 \end{bmatrix} - \begin{bmatrix} 6 \\ 90 \end{bmatrix} \right) \left(\begin{bmatrix} 7 \\ 100 \end{bmatrix} - \begin{bmatrix} 6 \\ 90 \end{bmatrix} \right)' = \\ = 5 \begin{bmatrix} -1 \\ -10 \end{bmatrix} \begin{bmatrix} -1 & -10 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 10 \end{bmatrix} \begin{bmatrix} 1 & 10 \end{bmatrix} = \\ = 5 \begin{bmatrix} 1 & 10 \\ 10 & 100 \end{bmatrix} + 5 \begin{bmatrix} 1 & 10 \\ 10 & 100 \end{bmatrix} = \begin{bmatrix} 10 & 100 \\ 100 & 1000 \end{bmatrix}$$

Poszukiwaną statystyką wielowymiarową weryfikującą hipotezę zerową o braku wpływu konkretnego czynnika (lub współdziałania czynników) na zbiór p zmiennych objaśnianych łącznie, najbardziej zbliżoną do naszych intuicji, jest statystyka podobna do statystyki F . Niżej zaprezentowano szczegółową postać statystyk dla p -wymiarowej i dwuczynnikowej analizy wariancji, czyli dla najprostszego przypadku umożliwiającego testowanie efektów interakcyjnych:

– hipoteza $H_0: \alpha_i = \mathbf{0}$ dla $i = 1, 2, \dots, k$

$$(12) \quad \tilde{F} = \frac{f_R - p + 1}{f_A p} \text{tr}(H_A G^{-1})$$

ze stopniami swobody g_1 i g_2 (różnymi od liczb f_A i f_R , będącymi stopniami swobody odpowiednio macierzy H_A i G)⁴;

– hipoteza $H_0: \beta_j = \mathbf{0}$ dla $j = 1, 2, \dots, w$

$$(13) \quad \tilde{F} = \frac{f_R - p + 1}{f_B p} \text{tr}(H_B G^{-1})$$

ze stopniami swobody g_1 i g_2 , przy czym jeśli g_2 – stopnie swobody mianownika – są identyczne dla wszystkich trzech testów \tilde{F} , to g_1 jest z reguły inną liczbą niż wyżej wprowadzone g_1 (które zależy od liczby poziomów czynnika A, podczas gdy teraz wprowadzone g_1 zależy od liczby poziomów czynnika B);

– hipoteza $H_0: (\alpha\beta)_{ij} = \mathbf{0}$ dla $i = 1, 2, \dots, k$ oraz $j = 1, 2, \dots, w$

$$(14) \quad \tilde{F} = \frac{f_R - p + 1}{f_{A \times B} p} \text{tr}(H_{A \times B} G^{-1})$$

ze stopniami swobody g_1 (zależnymi od liczby poziomów przecięć czynników) i g_2 .

Kryterium śladowym albo statystyką T^2 -Hotellinga jest statystyka postaci:

$$(15) \quad T^2 = tr(HG^{-1})$$

Rozkład z próby tej statystyki można aproksymować za pomocą rozkładu prawdopodobieństwa F Fishera–Snedecora po wyznaczeniu stopni swobody f_1 i f_2 (we wzorach (12), (13) i (14) stopniom tym odpowiadają odpowiednio f_A , f_B i $f_{A \times B}$ dla f_1 oraz f_R dla f_2). Dla całej statystyki testowej \tilde{F} przy prawdziwości hipotezy zerowej wyznacza się stopnie swobody g_1 i g_2 , przy których \tilde{F} ma w przybliżeniu rozkład F Fishera–Snedecora. (dla zaznaczenia tego przybliżenia nad literą F znajduje się wężyk).

Zweryfikujemy prawdziwość hipotezy zerowej dla analizowanego przykładu. Z (12) wynika, że – poza macierzą \mathbf{H}_A – musimy obliczyć \mathbf{G}^{-1} (odwrotność macierzy \mathbf{G}):

$$(16) \quad G = \begin{bmatrix} 8 & 130 \\ 130 & 3200 \end{bmatrix}$$

(wyznacznik macierzy G jest różny od zera, istnieje zatem macierz odwrotna)

$$(17) \quad G^{-1} = \begin{bmatrix} \frac{3200}{8700} & \frac{-130}{8700} \\ \frac{-130}{8700} & \frac{8}{8700} \end{bmatrix} = \begin{bmatrix} 0,36782 & -0,01494 \\ -0,01494 & 0,00092 \end{bmatrix}$$

Podobnie jak \mathbf{G} , macierz \mathbf{G}^{-1} jest macierzą symetryczną:

$$\begin{aligned} tr(HG^{-1}) &= tr\left(\begin{bmatrix} 10 & 100 \\ 100 & 1000 \end{bmatrix} \begin{bmatrix} 0,36782 & -0,01494 \\ -0,01494 & 0,00092 \end{bmatrix}\right) = \\ &= tr\begin{bmatrix} 2,183908 & \bullet \\ \bullet & -0,574713 \end{bmatrix} = 1,609195 \end{aligned}$$

Macierz \mathbf{HG}^{-1} jest macierzą symetryczną, więc w miejscu kropek powinna być taka sama liczba. Ponieważ potrzebne są wyłącznie elementy głównej przekątnej tej macierzy, liczba ta nie została wyznaczona.

W konsekwencji

$$(18) \quad \tilde{F} = \frac{f_R - p + 1}{f_H p} tr(HG^{-1}) = \\ = \frac{8 - 2 + 1}{1 \cdot 2} 1,609195 = 5,632184$$

Ponieważ dla analizy jednoczynnikowej stopnie swobody g_1 i g_2 wynoszą odpowiednio (por. Ahrens i Läuter, 1979):

$$(19) \quad g_1 = \begin{cases} \frac{(k-1)p(n-k-p)}{n-(k-1)p-2}, \\ \infty, \end{cases} \\ \text{gd}y \quad n-(k-1)p-2 > 0, \\ \text{gd}y \quad n-(k-1)p-2 \leq 0, \\ g_2 = n-k-p+1$$

skąd $g_1 = 2$, a $g_2 = 7$ (gdź k odpowiada liczbie poziomów czynnika, czyli 2, $n = 10$, a $p = 2$). Wartość krytyczna dla poziomu istotności $\alpha = 0,05$ wynosi 4,74. Zatem hipotezę zerową o braku efektów działania czynnika A na strukturę zmiennych Y_1 i Y_2 należy odrzucić.

Dla sytuacji III naszego przykładu, w której skorelowanie zmiennych objaśniających jest silne i relatywnie najwyższe, otrzymana wartość statystyki $\tilde{F}_{III} = 5,6322$. Dla zaspokojenia ciekawości odnotujemy w tym miejscu, że dla sytuacji I oraz II naszego przykładu wartości statystyki testu wynoszą odpowiednio $\tilde{F}_I = 9,1304$, przy identycznych jak wyżej g_1 i g_2 jest to wartość z obszaru krytycznego, natomiast $\tilde{F}_{II} = 4,3750$ i jest to wartość mniejsza od wartości krytycznej (przy stopniach swobody jak wyżej). Wynik ten jest zgodny z naszą wcześniejszą intuicją, jako że oczekiwano wyższych wartości statystyki weryfikującej hipotezę o braku wpływu czynnika w sytuacji I i III oraz niższej wartości tej statystyki w sytuacji II.

Z formalnego punktu widzenia przedstawione powyżej rozwiązanie nie jest jedynym możliwym. Teoretycznie dla określonych założeń metody istnieje nieskończenie wiele dopuszczalnych rozwiązań, czyli możliwych postaci statystyk weryfikujących hipotezy zerowe takie jak wyżej. Zanim zaprezentujemy ważniejsze z tych funkcji, mające pożądane własności formalne, niezbędna jest pewna dygresja.

Każda macierz kwadratowa \mathbf{A} , mająca tyle samo kolumn co wierszy, np. p , może być interpretowana jako pewien sposób przekształcenia p -wymiarowych wektorów. Każdemu wektorowi z takiej p -wymiarowej przestrzeni, np. \mathbf{v} , można przyporządkować wektor \mathbf{x} , równy $\mathbf{A} \cdot \mathbf{v}$, czyli uzyskany z pomnożenia wektora pierwotnego przez macierz \mathbf{A} . Taka operacja powoduje na ogół zmianę długości wektora (który może się skrócić lub wydłużyć), ale – co ważniejsze – wektor z reguły zmieni swój kierunek w przestrzeni. Jednakże są takie wektory w przestrzeni (np. w naszej przestrzeni obserwacji), które zmieniają swoją długość, ale nie zmieniają orientacji przestrzennej, tzn. wskazują ten sam kierunek, o czym w języku matematycznym mówimy, że wektor \mathbf{x} jest równoległy do \mathbf{v} . Takie wyjątki nazywane są wektorami własnymi macierzy \mathbf{A} . Innymi słowy, istnieje taka liczba λ , której pomno-

zenie przez wektor przyniesie ten sam efekt, co pomnożenie macierzy \mathbf{A} przez ten wektor, czyli $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$. Jeśli liczba ta jest różna od jedności, zmieni się tylko długość tego wektora. Jeśli jest ona równa jedności, nawet długość wektora się nie zmienia. Takie liczby λ nazywane są wartościami własnymi macierzy \mathbf{A} . A zatem wynik działania macierzy na wektor własny sprowadza się do pomnożenia go przez liczbę λ . Jeśli macierz \mathbf{A} jest nie tylko kwadratowa, ale także symetryczna, to wektory własne – odpowiadające różnym wartościom własnym – są ortogonalne.

Macierzą charakteryzującą uzyskane z badania dane, kwadratową i symetryczną, jest macierz $\mathbf{G}^{-1} \cdot \mathbf{H}$. Gdyby udało się wyznaczyć wartości i wektory własne tej macierzy (o p kolumnach i p wierszach), w przestrzeni wektorów obserwacji można byłoby wyznaczyć nowy układ współrzędnych z osiami prostopadłymi. Oznacza to opisanie badanego zjawiska nowymi zmiennymi w liczbie mniejszej albo co najmniej równej p .

Jest to atrakcyjna perspektywa, ponieważ dzięki rozwiązaniu zagadnienia własnego macierzy $\mathbf{G}^{-1} \cdot \mathbf{H}$ prowadzi do redukcji zbioru zmiennych początkowych. Wielowymiarowa analiza wariancji umożliwia nie tylko odpowiedź na pytanie, jak czynniki czy ich interakcje zmieniają wartości zmiennych objaśnianych, ale też wskazuje, które z tych zmiennych można wyeliminować jako redundantne, gdyż ich obecność w strukturze nie dostarcza nowych informacji o naturze wpływu czynników.

Zagadnienie wyznaczania nowych cech i eliminowania redundantnych jest rozwiązywane przez analizę dyskryminacji. Analiza ta ze zrozumiałych względów powinna być dokonywana równoległe z wielowymiarową analizą wariancji. A podstawowym argumentem matematycznym jest to, że kryterium śladowe T^2 [por. (15)] jest sumą wszystkich niezerowych wartości własnych macierzy $\mathbf{H} \cdot \mathbf{G}^{-1}$:

$$(20) \quad T^2 = \text{tr}(\mathbf{H}\mathbf{G}^{-1}) = \sum_{h=1}^t \lambda_h$$

gdzie $t \leq p$ oznacza liczbę niezerowych wartości własnych (a tym samym liczbę wektorów własnych),

$$(21) \quad t = \min(p, s)$$

s – liczba liniowo niezależnych kolumn (wierszy) macierzy \mathbf{H} .

Liczbę t można wyznaczyć bardzo łatwo. Pamiętając, że macierz \mathbf{H} zawsze dotyczy licznika utworzonego z wariancji średnich z poziomów czynników albo licznika utworzonego z wariancji interakcji, s równe jest liczbie poziomów odpowiedniego czynnika – 1 lub dla interak-

cji – iloczynowi liczby poziomów obydwu czynników pomniejszonych o 1. Liczba s odpowiada liczbie stopni swobody dla wariancji jednowymiarowej.

Z postulatów formalnych dotyczących postaci testu weryfikującego H_0 wynika, że postać statystyki testu wielowymiarowego musi być funkcją wartości własnych macierzy $\mathbf{H} \cdot \mathbf{G}^{-1}$. Jedną z dopuszczalnych statystyk (funkcji) jest \tilde{F} , będąca – jak wynika z (20) – funkcją sumy wartości własnych macierzy $\mathbf{H} \cdot \mathbf{G}^{-1}$. Innymi dopuszczalnymi statystykami oprócz kryterium śladowego Lawleya-Hotellinga (por. Ahrens i Läuter, 1979; Tatsuoka, 1971; Timm, 1975) są:

– kryterium Wilksa

$$(22) \quad \Lambda = \frac{|\mathbf{G}|}{|\mathbf{H} + \mathbf{G}|} = \prod_{h=1}^t \frac{1}{1 + \lambda_h}$$

– kryterium Pillaia

$$(23) \quad V = \sum_{h=1}^t \left(1 - \frac{1}{1 + \lambda_h} \right)$$

– kryterium Roya maksymalnej wartości własnej, λ_1 , lub w nieco innej postaci:

$$(24) \quad \frac{\lambda_1}{1 + \lambda_1}$$

Warto zauważyć, że dla przypadku dwóch analizowanych grup, kryterium śladowe T^2 Lawleya-Hotellinga może być traktowane jako wielowymiarowy analog statystyki t testu t -Studenta dla dwóch niezależnych populacji.

Inną względem \tilde{F} postacią statystyki dla analizy jednoczynnikowej, opierającą się na kryterium Λ -Wilksa, jest:

$$(25) \quad F = \frac{1 - \Lambda^{\frac{1}{k}}}{\Lambda^{\frac{1}{k}}} \cdot \frac{k(2f_R + f_H - p - 1) - pf_H + 2}{2pf_H}$$

gdzie

$$(26) \quad k = \sqrt{(p^2 f_H^2 - 4) / (p^2 + f_H^2 - 5)}$$

przy stopniach swobody g_1 i g_2 :

$$(27) \quad g_1 = pf_H \quad g_2 = \frac{1}{2}(k(2f_R + f_H - p - 1) - pf_H + 2)$$

(gdy stopnie swobody macierzy \mathbf{H} wynoszą $f_H = 1$, wtedy przyjmuje się także $k = 1$ przy $p = 2$).

Kryterium Λ Wilksa można rozważać jako wielowymiarowe uogólnienie statystyki F w jednowymiarowej analizie wariancji ANOVA, a dla przypadku jednowymiarowego wartość Λ jest odwrotnie proporcjonalna do wartości statystyki F :

$$(28) \quad \Lambda_{p=1} = \frac{1}{1 + [(k-1)/(n-k)]F}$$

Jako że prawa strona równania stanowi monotoniczną funkcję statystyki F , oznacza to, że testowanie wielowymiarowej hipotezy zerowej dla przypadku jednowymiarowego sprowadza się do zastosowania statystyki F (Tatsuoka, 1971).

W przypadku dwóch analizowanych grup dla dowolnej liczby zmiennych objaśnianych zależności pomiędzy wartościami kryterium śladowym T^2 Lawleya-Hotellinga i kryterium Λ Wilksa można przedstawić w następujący sposób (Johnson i Wichern, 2002):

$$(29) \quad \frac{1}{\Lambda} = \left(\frac{T^2}{n-2} \right) - 1$$

Im większa wartość współczynnika T^2 , tym mniejsza wartość współczynnika Λ .

Dla sytuacji III analizowanego przykładu wartość statystyki F pokrywa się z wartością statystyki \tilde{F} :

$$(30) \quad \Lambda = \frac{1}{1 + 1,609195} = 0,38326$$

(gdyż dysponujemy zaledwie jedną wartością własną na podstawie (21))

$$(31) \quad f_H = 1, f_R = 8, p = 2$$

stąd $k = 1$, zatem

$$(32) \quad F = \frac{1 - 0,38326}{0,38326} \cdot \frac{14}{4} = 5,6322$$

przy $g_1 = 2$ i $g_2 = 7$

W Tabeli 2 przedstawiono wyniki testu Boxa oraz wielowymiarowego testu F dla wszystkich sytuacji, uzyskane w pakiecie SPSS.

Równoważność wartości statystyki F i \tilde{F} , jak wspomniano wyżej, nie jest – niestety – zasadą ogólną. W bardziej skomplikowanych planach badawczych war-

tości obydwu statystyk mogą się różnić. Może się więc zdarzyć, że w przypadku F wynik jest istotny statystycznie, zaś dla \tilde{F} – nieistotny lub odwrotnie. Odpowiedź na pytanie, dlaczego tak jest – nie należy do łatwych i wiąże się z dwoma aspektami sprawy.

Pierwszym jest pytanie o moc każdego z wymienionych testów (a więc intuicyjnie o łatwość odrzucania hipotezy zerowej). Odpowiedź w tej sprawie nie jest optymistyczna. Badanie Gnanadesikana i innych w latach 60. XX w. mocy tych testów dla przypadku dwu cech objaśnianych przez model (por. Ahrens i Lauter, 1979, s. 63) ujawniły w zasadzie „przewagę” F nad \tilde{F} . W zasadzie, gdyż nie zawsze tak się zdarzało.

Drugim jest pytanie o odporność testów na pogwałcenie założeń metody. W 1981 roku Kariya (por. Zieliński, 1985) opublikował wyniki dużych badań symulacyjnych nad mocą testów parametrycznych w sytuacji „normalności” i „nienormalności” rozkładów zmiennych. Okazało się, iż w obu przypadkach jednostajnie najmocniejszym testem jest T^2 Lawleya-Hotellinga.

W literaturze brak zgody odnośnie do tego, która ze statystyk i w jakiej sytuacji stanowi najlepszy wybór. W większości przypadków odchylenia od wielowymiarowej normalności mają niewielki wpływ na rzeczywisty poziom prawdopodobieństwa popełnienia błędu pierwszego rodzaju (Everitt, 1979), jednakże moc testów znacznie spada w przypadku platykurtyczności rozkładu zmiennych objaśnianych (Olson, 1974). Zwykle, zarówno w przypadku małych, jak i dużych efektów zmiennych objaśniających, wartości wszystkich czterech statystyk prowadzą do podobnych decyzji. Jeśli różnice między grupami dotyczą tylko jednego wymiaru, najmocniejszym z rozwiązań jest statystyka oparta na kryterium Roya. Dla małych prób kryterium Pillai gwarantuje stosunkowo odporną statystykę na naruszenie założenia o równości macierzy kowariancji (Olson, 1976), a także w przypadku występowania współliniowości zmiennych objaśnianych. Jednakże, gdy założenie to nie jest spełnione, badacz nadal musi się liczyć z możliwością podjęcia błędnej decyzji, jako że w takiej sytuacji rzeczywiste w stosunku do nominalnego prawdopodobieństwo popełnienia błędu pierwszego rodzaju wzrasta, nawet gdy analizowane próby są równoliczne (Tang i Algina, 1999).

We wszystkich powyższych sytuacjach kryterium Wilksa nie ustępuje pod względem własności pozostającym, co oprócz faktu, że zostało wywiedzione w ramach metody największej wiarygodności (*maximum likelihood*, ML), jest najprawdopodobniej powodem jego największej popularności (Raykov i Marcoulides, 2008). Jednakże Λ Wilksa jest bardzo wrażliwa na występowanie przypadków odstających (*outliers*). Niedawno w literaturze

Tabela 2.

Wartości statystyki Boxa i F -wielowymiarowego

Sytuacja	Test Boxa		MANOVA	
	F	p	F	p
I	0,513	0,673	9,130	0,011
II	1,430	0,232	4,375	0,059
III	0,918	0,431	5,632	0,035

zaproponowano jej odporną postać opierającą się na estymatorze MCD (*minimum covariance determinant estimator*) dla przypadku jednoczynnikowej, wielowymiarowej analizy wariancji (Todorov i Filzmoser, 2010).

Sprawę dodatkowo komplikują siła i kierunek skorelowania zmiennych objaśniających, które wpływają na moc parametrycznych testów wielowymiarowych (Cole, Maxwell, Arvey i Salas, 1993b), dlatego coraz częściej w przypadku pogwałcenia założeń proponowane są rozwiązania nieparametryczne (Bathke, Harrar, Laurence i Madden, 2008; Finch, 2005; Finch i Davenport, 2009). Bez wątpienia założenia, jakich spełnienia wymaga wielowymiarowa analiza wariancji, ograniczają zakres jej aplikacji w badaniach społecznych. W obecnie rozwijanych alternatywnych podejściach statystyki testowe opierają się na odległościach pomiędzy obserwacjami i są porównywane z rozkładami permutacyjnymi. W metodach permutacyjnych (*Multiresponse permutation procedures* – MRPP) jako statystyka testowa wykorzystywana jest średnia (ważona) odległość wewnątrzgrupowa (Mielke i Berry, 2007). W wywodzącym się z ekologii podejściu, zaproponowanym przez McArdle i Anderson (2001), określanym jako permutacyjna (nieparametryczna) MANOVA, proponowana statystyka „pseudo-F” została skonstruowana na podstawie metody podziału wariancji macierzy niepodobieństwa (odległości).

Warto zauważyć, że w specyficznych przypadkach, gdy badacz zmienne mierzalne „silnie” skategoryzuje, tworząc zaledwie kilka klas pomiaru, oczywiście nie może posługiwać się MANOVA. Powinien zastosować analizę korespondencji, w której kategorii wartości wszystkich czynników stanowiłyby jeden zbiór zmiennych (tzw. zero-jedynkowych) natomiast kategorie zmiennych objaśnianych – drugi zbiór (por. Aranowska i Ciok, 1992).

Poszukuje się także lepszych rozwiązań parametrycznych, czego przykładem może być uogólniony test F (*generalized F-test*), szczególnie użyteczny w przypadku wielu zmiennych objaśnianych i małych prób (Liang i Tang, 2009). W literaturze nie rekomenduje się analizowania zbyt dużej liczby zmiennych objaśnianych. Moc testów wielowymiarowych maleje wraz ze wzrostem liczby zmiennych objaśnianych (DasGupta i Perlman, 1973). Sensownym rozwiązaniem ograniczającym liczbę zmiennych może być na przykład wcześniejsze wykorzystanie w tym celu analizy składowych głównych.

Większość programów statystycznej analizy danych podaje wartości wszystkich czterech statystyk i zwykle prowadzą one do tych samych decyzji odnośnie do hipotezy zerowej. Ze względu na swoje zalety, a także na tradycję (zostało wyprowadzone jako pierwsze) najpowszechniej wykorzystywane jest kryterium Wilksa. W wypadku roz-

bieżności decyzji podejmowanych co do hipotezy zerowej badacz musi poddać analizie macierze kowariancji i wartości własne oraz – w świetle diskutowanych wyżej ograniczeń każdej ze statystyk – ocenić sprzeczne konkluzje.

Wspomniano wyżej, że wyznaczanie wartości własnych macierzy $\mathbf{H} \cdot \mathbf{G}^{-1}$ niezbędne jest nie tylko ze względu na potrzebę wyznaczania wartości wielowymiarowego testu F czy \tilde{F} , lecz także dlatego, iż wiąże się to z możliwością redukcji zbioru zmiennych objaśnianych do zbioru mniejszego, dla którego wartość testu w dalszym ciągu jest istotna. Innymi słowy, chodzi o wyznaczenie takiej minimalnej w sensie liczby zmiennych struktury cech, na którą czynniki wpływają najsilniej. Metodą wskazującą najsilniejszą dyskryminację wpływów poziomów czynników na badane zmiennej jest analiza dyskryminacji.

Zastosowanie metody MANOVA do analizy efektywności pracy menedżerów

Celem przywołanych niżej badań (Strzałecki i Tomaszewicz, 2004), „była próba wyjaśnienia psychologicznych mechanizmów efektywności pracy menedżerów i osiągania przez nich sukcesu, rozumianego jako wynik twórczego zarządzania przez nich ludźmi i zasobami, w tym również informacyjnymi” (s. 127). Zakładano, że zarządzający efektywnie wykazują wobec ludzi niedziałających w biznesie specyficzną konfigurację cech ludzi twórczych opisaną przez model Stylu Twórczego Zachowania (STZ) Strzałeckiego (2003), częściowo operacjonalizowaną za pomocą kwestionariusza Style Zachowania Się. Ze względu na temperamentalne uwarunkowania efektywności działań menedżerskich, do badań włączono kwestionariusz temperamentu FCZ-KT. Badanie dziedziny intelektualnej ograniczono do odtworzenia adaptacyjnej giętkości myślenia (według modelu SIM) mierzonej na skali Usuwania Zapalek Guilforda, natomiast aspekty aksjologiczne zawężono, uwzględniając konstrukt koherencji Antonovsky’ego operacjonalizowany Kwestionariuszem Poczucia Koherencji (SOC).

W przeprowadzonym badaniu podstawowym problemem badawczym było zatem określenie różnic dotyczących temperamentalnych, osobowościowych oraz poznawczych aspektów funkcjonowania menedżerów wysokiego szczebla (dyrektorzy i prezesi firm) oraz pracowników sfery budżetowej. Badano grupy o tej samej liczebności ($n_1 = n_2 = 43$). Analizę wyników przeprowadzono za pomocą jednoczynnikowej, 15-wymiarowej analizy wariancji (MANOVA). Do analizy włączono wyniki uzyskane przez osoby badane z obu grup we wszystkich narzędziach (6 skal kwestionariusza temperamentu FCZ-KT + 5 skal Kwestionariusza Style Zachowania

Się + 3 skale Kwestionariusza Poczucia Koherencji Antonovsky'ego + 1 skala Testu Usuwania Zapalek Guilforda = 15). Wynik ogólny kwestionariusza Poczucia Koherencji Antonovsky'ego (SOC), jako że stanowi on sumę wyników trzech skal tego kwestionariusza, został wyłączony z analizy.

Podstawowym założeniem dla stosowania modelu MANOVA jest założenie o równości macierzy kowariancji w porównywanych próbach weryfikowane za pomocą testu Boxa. W przypadku uzyskanych wyników założenie to zostało spełnione (wartość statystyki F opierającej się na statystyce M testu Boxa była równa 1,179; $p = 0,088 > 0,05$), co pozwoliło na zastosowanie modelu do analizowanych danych.

Wartość wielowymiarowego testu F opierającego się na statystyce lambda Wilksa, podobnie jak wszystkich pozostałych statystyk, wyniosła 4,607; $p = 0,0005$ i okazała się istotna statystycznie na przyjętym poziomie istotności $\alpha = 0,05$. Okazało się zatem, że osoby aktywnie zarządzające na co dzień wobec osób niepodających

działalności w biznesie różnią się nasileniem analizowanej struktury zmiennych. Profesjonalizm menedżerski wyjaśnia około 50,0% zmienności analizowanej struktury cech ($\eta^2_{\text{wielowymiarowe}} = 0,497$).

Wartości statystyki t testu t -Studenta dla poszczególnych skal wykorzystanych w badaniu narzędzi, przedstawione zostały w Tabeli 3. W tabeli zamieszczono też dodatkowo wartość statystyki t otrzymaną dla porównania – wyłączonych z analizy wielowymiarowej – wyników ogólnych kwestionariusza Poczucia Koherencji Antonovsky'ego (SOC). Tabela zawiera ponadto oszacowania wielkości efektów działania czynnika dla poszczególnych zmiennych przeprowadzone za pomocą współczynnika d -Cohena (Cohen, 1977; 1992).

Wartości jednowymiarowych statystyk F , które z reguły podaje się po wielowymiarowej analizie wariancji, w przypadku dwóch populacji niezależnych równe są t^2 . Co zatem brać pod uwagę, rozważając wyniki badań: t czy F ? Za wyborem statystyki t przemawiają następujące argumenty. Po pierwsze, test t umożliwia weryfi-

Tabela 3.
Wartości statystyki t

Nazwa skali	Grupa kryterialna		Grupa kontrolna		t	p	d
	M	SD	M	SD			
FCZ-KT							
Żwawość	17,28	2,66	15,44	3,79	2,61	0,005	0,57
Perseweratywność	10,19	4,81	11,86	4,14	1,73	0,040	0,38
Wrażliwość sensoryczna	14,74	4,20	15,12	3,52	0,44	nist.	0,10
Reaktywność emocjonalna	7,00	4,40	11,16	5,64	3,82	0,0003	0,83
Wytrzymałość	11,42	5,11	9,56	4,24	1,84	0,035	0,40
Aktywność	12,16	4,92	7,60	4,05	4,69	0,0003	1,02
Style Zachowania Się							
Aprobata życia	22,88	4,14	20,53	3,89	2,71	0,004	0,59
Silne ego	37,93	9,19	31,28	11,00	3,04	0,0015	0,66
Samorealizacja	34,70	8,54	28,91	7,01	3,44	0,0005	0,75
Giętkość procesów poznawczych	45,84	9,26	40,93	9,66	2,40	0,009	0,53
Wewnętrzna sterowność	37,95	7,09	32,93	8,81	2,91	0,0025	0,64
SOC							
Poczucie zrozumiałości	51,16	7,15	44,47	9,02	3,82	0,0003	0,83
Poczucie zaradności	57,23	5,45	49,09	8,37	5,34	0,0003	1,17
Poczucie sensowności	48,14	5,04	41,81	6,13	5,22	0,0003	1,14
Wynik ogólny SOC	156,53	14,63	135,37	20,42	5,53	0,0003	1,21
Test Usuwania Zapalek							
Adaptacyjna giętkość myślenia	37,65	10,52	23,37	9,68	6,56	0,0003	1,43

p – dla testu jednostronnego

kowanie hipotezy kierunkowej, zatem przy tym samym poziomie istotności α minimalizuje prawdopodobieństwo popełnienia błędu II rodzaju (β), tym samym jest testem mocniejszym. Po drugie, na poziomie treściowym ujawnia więcej cech dyskryminowanych przez poziomy czynnik. Tak było w tym przypadku. Jednowymiarowy test F nie ujawnił różnic poziomów perseweratywności ($p = 0,08$) i wytrzymałości ($p = 0,07$), które ujawnił test t -Studenta. Wielkości efektów dla tych zmiennych, wynoszące odpowiednio 0,38 i 0,40, przekraczają wielkość uznawaną za małą ($d = 0,20$), zbliżając się do przeciętnej wartości efektu ($d = 0,50$).

Następnym niezbędnym etapem analizy danych jest przeprowadzenie analizy dyskryminacyjnej w celu wyodrębnienia skal, na które czynnik (przynależność do grupy) wpływa najsilniej. Analizą objęto wyniki 15 skal czterech wykorzystanych w badaniu kwestionariuszy, wyłączając z niej – rzecz jasna – wynik ogólny kwestionariusza Poczucia Koherencji (SOC).

Wielokrotnie stosowane – dla każdej zmiennej oddzielnie – jednowymiarowe i odpowiednio czynnikowe analizy wariancji (ANOVA) traktowane jako kontrasty parametrów pokażą tylko, jaki jest samodzielny wkład każdej zmiennej do różnicy między grupami. Natomiast analiza dyskryminacyjna odtworzy cząstkowy wkład każdej zmiennej z grupy łącznie analizowanych zmiennych związanych konkretną siecią wartości współczynników r -Pearsona, czyli konkretnie wzajemnie się determinujących. Komplementarne stosowanie obydwu metod polecane jest w literaturze statystycznej już od lat 70. XX w. (Bergen i Seling, 1978; Enders, 2003). W pakietach statystycznych starszej generacji (np. OSIRIS czy MVD) analiza dyskryminacyjna jako procedura uzupełniająca wielowymiarową analizę wariancji była przeprowadzana standardowo.

Metodę analizy dyskryminacyjnej można porównać z metodą regresji wielokrotnej, a zatem z sytuacją przewidywania wartości jednej zmiennej (objaśnianej) na podstawie zbioru innych zmiennych (objaśniających). Rezultatem jest równanie regresji, za pomocą którego możemy przewidywać wartości zmiennej objaśnianej. W modelu regresji wielokrotnej wszystkie zmienne (objaśniana i objaśniające) powinny być mierzalne. W modelu analizy dyskryminacyjnej zmienna objaśniana jest traktowana jako nominalna (konstituują ją poziomy czynnik, czyli zmiennej objaśniającej), a rezultatem analizy jest postać funkcji dyskryminacyjnej pozwalającej na trafne przewidywanie przynależności do określonej grupy. Analiza umożliwia zatem wyodrębnienie zmiennych najsilniej różnicujących badane próby, a interpretacja wag dyskryminacyjnych dla każdej zmiennej jest podobna do

tej, jaką stosujemy przy interpretacji wag beta w analizie regresji wielokrotnej. Wielkość poszczególnych wag dla każdej zmiennej stanowi informację o udziale każdej z tych zmiennych w efektywności różnicowania pomiędzy interesującymi nas grupami. Warto zauważyć, że dla przypadku dwugrupowego (podobnego do tu analizowanego) obydwie metody generują identyczne rozwiązania, a wartości wag dyskryminacyjnych będą proporcjonalne do wartości cząstkowych współczynników w równaniu regresji wielokrotnej.

Inną wygodną miarą siły dyskryminacyjnej poszczególnych zmiennych jest współczynnik korelacji każdej zmiennej z wartościami funkcji dyskryminacyjnej. Kwadrat tego współczynnika, choć oznacza stopień wyjaśniania zmienności danej zmiennej przez wartości utworzonej funkcji dyskryminacyjnej, to z racji jego symetryczności może służyć do określenia zakresu determinacji wartości funkcji dyskryminacyjnej przez poszczególne uwzględnione w analizie zmienne (por. Ahrens i Läuter, 1979).

W Tabeli 4 zostały przedstawione wartości wag dyskryminacyjnych, czyli standaryzowanych wartości współczynników funkcji dyskryminacyjnej dla analizowanych 15 skal czterech kwestionariuszy. Są to współczynniki standaryzowane (niezależne od jednostek, w których były mierzone zmienne), możemy zatem porównywać ich wartości. Tabela zawiera także wartości współczynników korelacji każdej zmiennej z wartościami funkcji dyskryminacyjnej.

Otrzymana funkcja dyskryminacyjna pozwala na istotne statystycznie różnicowanie pomiędzy średnimi punktów dyskryminacyjnych w obu grupach (wartość lambdy Wilksa = 0,503; $p = 0,0005$). Uzyskana wartość współczynnika korelacji kanonicznej jest stosunkowo wysoka, równa 0,705. Funkcja charakteryzuje się zadowalającą efektywnością różnicowania. Umożliwia ona poprawne zaklasyfikowanie 87,2% przypadków do obu grup: menedżerów i pracowników budżetowych. Rezultaty klasyfikacji przedstawione zostały w Tabeli 5.

Efektywność różnicowania na podstawie funkcji dyskryminacyjnej jest wyższa niż obliczona na podstawie kryterium proporcjonalnych szans (C). Kryterium to wyznacza się w następujący sposób: $C = p^2 + (1 - p)^2$, gdzie p jest proporcją przypadków w jednej grupie (Aczel, 2000, s. 893). Dla analizowanych danych wynosi ono 0,5, czyli 50%. Współczynnik trafności różnicowania otrzymanej funkcji dyskryminacyjnej znacznie przekracza tę wartość ($39 + 36 = 75$ poprawnie zaklasyfikowanych przypadków na wszystkie 86 przypadków, czyli 87,2%). Łatwo sprawdzić, że proporcja poprawnie zakwalifikowanych osób (0,872) istotnie statystycznie różni się od pro-

Tabela 4.

Wartości standaryzowanych współczynników funkcji dyskryminacyjnej oraz współczynników korelacji cech pierwotnych z cechą dyskryminacyjną

Nazwa skali	Wagi dyskryminacyjne	Współczynniki korelacji
FCZ-KT		
Żwawość	-0,081	0,286
Perseweratywność	0,355	-0,190
Wrażliwość sensoryczna	-0,064	-0,049
Reaktywność emocjonalna	-0,445	-0,419
Wytrzymałość	-0,348	0,202
Aktywność	0,265	0,515
Style Zachowania się		
Aprobata życia	0,016	0,298
Silne ego	0,006	0,334
Samorealizacja	0,190	0,377
Giętkość procesów poznawczych	0,127	0,264
Wewnętrzna sterowność	-0,387	0,320
SOC		
Poczucie zrozumiałości	0,139	0,419
Poczucie zaradności	0,067	0,587
Poczucie sensowności	0,465	0,574
Test Usuwania Zapalek		
Adaptacyjna giętkość myślenia	0,669	0,720
Wartość λ Wilksa	0,503; $p = 0,0005$	
Wartość współczynnika korelacji kanonicznej	0,705	

Tabela 5.

Rezultaty klasyfikacji dokonanej na podstawie funkcji dyskryminacyjnej

Rzeczywista przynależność do grupy	Przewidywana przynależność do grupy	
	Menedżerowie	Pracownicy budżetowi
Menedżerowie	39	4
Pracownicy budżetowi	7	36

porcji osób zakwalifikowanych poprawnie przypadkowo (0,50), wykorzystując w tym celu statystykę t postaci $t = (P - 0,5) / \sqrt{(0,5)(1 - 0,5) / n}$, gdzie P – oznacza proporcję poprawnie zakwalifikowanych osób, a $n = n_1 + n_2$ (por. Hair i in., 1995). Otrzymana wartość statystyki t równa jest 6,70 i przy poziomie istotności $\alpha = 0,05$ przekracza wartość krytyczną dla liczby stopni swobody równej $n - 2 = 86 - 2 = 84$ wynoszącą odpowiednio 1,666 (jednostronnie).

Na podstawie analizy Tabeli 4, podobnie jak na podstawie analizy wartości statystyki t i d (por. Tabela 3) możemy stwierdzić, że zmienna adaptacyjna giętkość myślenia (mierzona testem Usuwania Zapalek Guilforda) wnosi największy udział do różnicowania między grupami. Otrzymana dla niej wartość wagi dyskryminacyjnej (0,669), jak i wartość współczynnika korelacji (0,720), jest najwyższa. Jednakże, jak widać, uszeregowanie zmiennych ze względu na ich użyteczność w przewidywaniu przynależności do grupy uzyskane na podstawie wielkości standaryzowanych wag oraz wartości współczynników korelacji wykazuje pewne rozbieżności (por. Tabela 4). Na przykład, biorąc pod uwagę wartości wag dyskryminacyjnych, udział zmiennej Poczucie zaradności w różnicowanie między grupami (waga = 0,067) jest praktycznie żaden. Natomiast biorąc pod uwagę wartość jej korelacji z funkcją dyskryminacyjną, uplasowała się ona na drugim miejscu ($r = 0,587$). Co jest przyczyną takiego stanu rzeczy? Otóż trzy rodzaje informacji używane do oceny względnego znaczenia analizowanych zmiennych dla przewidywania przynależności grupowej (wartości statystyki F , standaryzowanych wag oraz współczynników korelacji) będą prowadziły do tych samych wniosków na temat oceny znaczenia zmiennych tylko wtedy, gdy skorelowanie między zmiennymi jest stosunkowo niewielkie. Kiedy pomiędzy zmiennymi występuje wysoki stopień skorelowania, ich uszeregowanie nie musi być takie samo. W takiej sytuacji wartości współczynników funkcji dyskryminacyjnej muszą być interpretowane ze znacznie większą ostrożnością. Podobnie jak w przypadku analizy regresji, mała standaryzowana waga może oznaczać nieistotność danej zmiennej dla rozróżnienia między grupami, ale może oznaczać także usunięcie jej wpływu ze względu na wysoki stopień współliniowości danych. We wskazywanym powyżej przypadku analiza wartości współczynników korelacji r -Pearsona wyznaczonych dla wszystkich par zmiennych ujawniła wysoką, dodatnią zależność pomiędzy zmiennymi Poczucie sensowności i Poczucie zaradności. Wartość współczynnika korelacji r -Pearsona pomiędzy tymi zmiennymi dla całej próby była równa 0,714; $p = 0,0005$, co oznacza, że

zmiennie podzielają 51% zmienności. Wartości te dla obu prób wynoszą odpowiednio: $r = 0,527$; $p = 0,0005$ dla menedżerów oraz $r = 0,674$; $p = 0,0005$ dla pracowników sfery budżetowej. Można podejrzewać, że na relatywnie wysoką korelację funkcji dyskryminacyjnej ze zmienną Poczucie zaradności (PZ) wpływa jej wysoki stopień skorelowania ze zmienną Poczucie sensowności (PS) oraz różny stopień skorelowania obu tych zmiennych ze zmienną najsilniej dyskryminującą obie populacje, czyli Adaptacyjną giętkością myślenia (AGM). Na Rysunku 4 przedstawione zostały wartości współczynników korelacji omawianych zmiennych w obydwu grupach.

Jak widać na Rysunku 4, zmiennie Poczucie sensowności i Adaptacyjna giętkość myślenia nie wykazują – na podstawie omawianych tu badań – istotnej relacji liniowej, natomiast taką relację sugerują związki Poczucia zaradności z Adaptacyjną giętkością myślenia oraz – co podkreślano wcześniej – Poczucia zaradności z Poczuciem sensowności. Zarówno MANOVA, jak i analiza dyskryminacyjna premią niezależność (liniową) zmiennych przy poszukiwaniu różnic w rozkładach cech w różnych populacjach (por. rozważania drugiego punktu niniejszego tekstu), dlatego w rozwiązaniu analizy dyskryminacyjnej pojawiła się niska wartość standaryzowanej wagi zmiennej Poczucie zaradności.

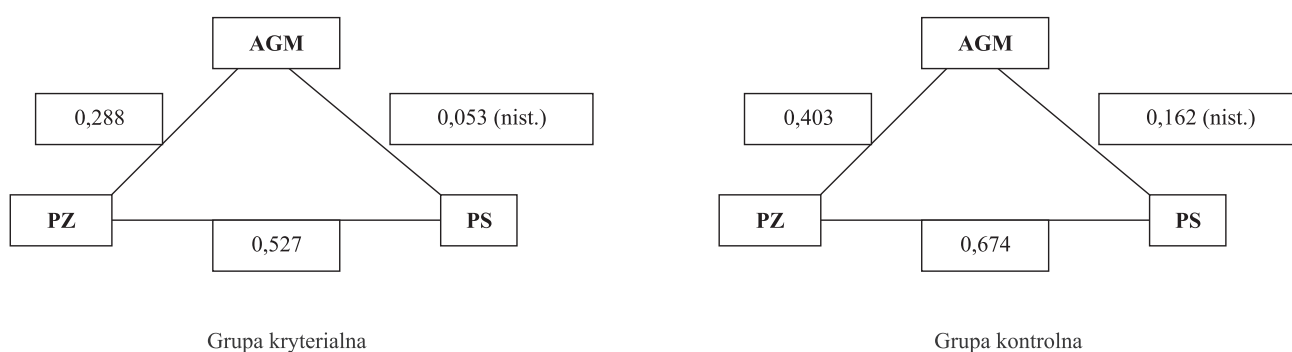
Wyznaczanie znaków wag cech dyskryminacyjnych ma zasadnicze znaczenie dla interpretacji treściowej wyników. Jest to chyba najbardziej znaczący dla badacza argument na rzecz stosowania metod wielowymiarowych. Wróćmy do przykładu. Mimo iż w zmiennych Żwawość, Wytrzymałość i Wewnętrzna sterowność wartości średnie menedżerów (grupy kryterialnej) były istotnie wyższe niż w grupie kontrolnej – podobnie jak w dwóch najsilniej różnicujących grupy zmiennych: Adaptacyjna giętkość

myślenia i Poczucie sensowności (por. Tabela 3) – ujemna waga tych trzech zmiennych w cesze dyskryminacyjnej (por. Tabela 4) oznacza, że im wyższe wartości tych cech w grupach, tym większe podobieństwo tych grup, mniejsza dyskryminacja międzygrupowa (mniejsza wartość cechy dyskryminacyjnej formalnie będącej wielokrotnym równaniem regresji). Zmienne Żwawość, Wytrzymałość i Wewnętrzna sterowność, gdy realizują się na poziomie wysokich wartości (skali), obniżają różnice międzygrupowe, co oznacza zwiększanie się prawdopodobieństwa zdarzenia polegającego na złym klasyfikowaniu takich osób do grup, nierozróżniania menedżerów od pracowników sfery budżetowej. Odwrotnie – zmienna Perseweratywność przy swoich rosnących wartościach zwiększa te różnice (por. Tabele 3 i 4). Takich implikacji badacz nie ma możliwości wykazać dzięki analizom jednowymiarowym.

Warto podkreślić, że interpretacji wyników badacz dokonuje, analizując równocześnie wartości czterech współczynników wyznaczonych dla każdej cechy pierwotnej: t , d , r i waga dyskryminacyjna. Tylko wartości ostatniego z nich uwzględniają relację połączeń cech pierwotnych, dlatego powinny być promowane podczas uogólnień treściowych.

Spójrzmy ponownie na wartości wspomnianych wyżej współczynników dla obu zmiennych zawartych w Tabeli 6.

Wartości trzech pierwszych współczynników dla Poczucia zaradności są porównywalne z odpowiednimi współczynnikami otrzymanymi dla Poczucia sensowności, natomiast wartości wag dyskryminacyjnych obydwu zmiennych różnią się dramatycznie. Bazując przy interpretacji wyników na zauważonym podobieństwie, badacz nie może określić, która ze zmiennych wnosi większy



Rysunek 4.

Wartości współczynników korelacji r -Pearsona pomiędzy zmiennymi Poczucie zaradności, Poczucie sensowności i Adaptacyjna giętkość myślenia w obydwu grupach.

Tabela 6.

Wartości statystyki t , d -Cohena, współczynnika korelacji r z cechą dyskryminacyjną oraz wartości wag dyskryminacyjnych dla wybranych dwóch zmiennych: Poczucia zaradności i Poczucia sensowności

Zmienna	t	d	r	waga
Poczucie zaradności	5,34	1,17	0,587	0,067
Poczucie sensowności	5,22	1,14	0,574	0,465

udział treściowy w dyskryminację wyróżnionych grup. Dopiero wyniki analizy dyskryminacyjnej umożliwiają ocenę roli tych zmiennych w postulowanym modelu.

W świetle toczącej się w literaturze dyskusji nad użytecznością obu rodzajów wskaźników: korelacji pomiędzy zmiennymi pierwotnymi i funkcją dyskryminacyjną (określanymi często mianem wskaźników struktury) oraz standaryzowanymi wagami dyskryminacyjnymi najlepszym rozwiązaniem wydaje się rekomendacja, że w identyfikacji zmiennych najsilniej dyskryminujących populacje badacz winien jednak brać pod uwagę obydwie miary. Obydwa wskaźniki umożliwiają określenie wkładu, jaki wnoszą w różnicowanie między populacjami dana zmienna pierwotna, jednakże ich wartości dostarczają odmiennych rodzajów informacji. Wskaźniki struktury – jako współczynniki korelacji całkowitej – ujmują ów wkład w sposób bezwzględny, nie uwzględniając powiązań między zmiennymi. Natomiast standaryzowane wagi dyskryminacyjne umożliwiają jego określenie w sposób względny, wyznaczony siłą powiązań pomiędzy zmiennymi pierwotnymi. Oba wskaźniki mają wady i zalety, nie ma jednak żadnego merytorycznego powodu, dla którego badacz miałby rezygnować z uwzględniania którejkolwiek z nich (por. Stevens, 2002, s. 289).

Ze względu na rozbieżność wskazań otrzymanych na podstawie wartości obu miar (por. Tabela 4) odnośnie do wielkości wkładu analizowanych zmiennych w dyskryminację pomiędzy populacjami, na tym samym zbiorze 15 zmiennych przeprowadzono powtórnie analizę dyskryminacyjną. Wykorzystano w niej inne dopuszczalne rozwiązanie formalne odtwarzające funkcje dyskryminacyjne, nazwane metodą *stepwise* (krokową), podejmując próbę zredukowania liczby zmiennych, czyli wyznaczenia reguły klasyfikacyjnej opartej na mniejszej liczbie zmiennych, tak by otrzymane rozwiązanie charakteryzowało podobną efektywność różnicowania.

Metody krokowe (*stepwise*), choć powszechnie wykorzystywane, często stanowią przedmiot zdecydowanej krytyki. Znakomitej ilustracji dostarcza stwierdzenie Thompsona (2001, s. 86): „Uważam, że metody krokowe

(np. regresja krokowa, krokowa deskryptywna analiza dyskryminacyjna) są złe, zgubne, kiepskie, bezwartościowe i błędne. Dodatkowo nie lubię ich”. Badacze wskazują na trzy problemy związane z wykorzystaniem metodologii krokowej: niepoprawną liczbę stopni swobody, kapitalizację błędu próby (*sampling error*), co ogranicza możliwość generalizacji uzyskanych rezultatów, oraz to, że za pomocą tych metod często nie udaje się wyodrębnić zbioru najlepszych predyktorów określonej wielkości (Huberty, 1989, 1994; Snyder, 1991; Thompson, 1989, 1995). Jednakże, gdy celem analizy dyskryminacyjnej jest wyłonienie zmiennych najsilniej różnicujących grupy, a nie przewidywanie przynależności grupowej obiektów, metody krokowe stosowane w kontekście wielowymiarowej analizy wariacji jako jej uzupełnienie, mimo swoich ograniczeń, mogą być użyteczne (por. Huberty, 1994), co dokumentuje przedstawione poniżej rozwiązanie.

W Tabeli 7 zostały przedstawione wartości wag dyskryminacyjnych dla otrzymanego rozwiązania.

Otrzymana funkcja dyskryminacyjna, opierająca się na dwóch tylko zmiennych: Adaptacyjnej giętkości myślenia oraz Poczuciu sensowności, pozwala na istotne statystycznie różnicowanie pomiędzy średnimi punktów dyskryminacyjnych w obu grupach (wartość lambdy Wilksa = 0,569; $p = 0,0005$). Uzyskana wartość współczynnika korelacji kanonicznej, równa 0,656, jest także stosunkowo wysoka i tylko nieznacznie niższa (o 0,049) od uzyskanej dla poprzedniej funkcji (równiej 0,705). Otrzymana funkcja charakteryzuje się także zadowalającą, zaledwie o 7% niższą w porównaniu z poprzednią funkcją, efektywnością różnicowania. Umożliwia ona poprawne zaklasyfikowanie 80,2% przypadków do obu grup: menedżerów i pracowników sfery budżetowej. Także w tym przypadku proporcja poprawnie zakwalifikowanych osób (0,802) istotnie statystycznie różni się od proporcji osób przypadkowo zakwalifikowanych poprawnie ($t = 5,60$). Rezultaty klasyfikacji przedstawione zostały w Tabeli 8.

Tabela 7.

Wartości standaryzowanych współczynników funkcji dyskryminacyjnej

Nazwa skali	Wagi dyskryminacyjne	Współczynniki korelacji
Adaptacyjna giętkość myślenia	0,760	0,823
Poczucie sensowności	0,572	0,655
Wartość Λ Wilksa	0,569; $p = 0,0005$	
Wartość współczynnika korelacji kanonicznej	0,656	

Tabela 8.
Rezultaty klasyfikacji dokonanej na podstawie funkcji dyskryminacyjnej

Rzeczywista przynależność do grupy	Przewidywana przynależność do grupy	
	Menedżerowie	Pracownicy budżetowi
Menedżerowie	34	9
Pracownicy budżetowi	8	35

Wykorzystując niestandardyzowane wartości współczynników otrzymanej funkcji dyskryminacyjnej (wynoszące odpowiednio: 0,075 dla zmiennej Adaptacyjna giętkość myślenia i 0,102 dla zmiennej Poczucie sensowności) oraz wartości średnich obu zmiennych w obu grupach, możemy wyznaczyć średnie wartości różnicujące (średnie wartości funkcji dyskryminacyjnej) dla każdej z grup:

$$\bar{Y}_{\text{menedżerowie}} = 0,075 \cdot 37,67 (\text{średnia zmiennej AGM}) + 0,102 \cdot 48,14 (\text{średnia zmiennej PS}) = 7,736;$$

$$\bar{Y}_{\text{pracownicy budżetowi}} = 0,075 \cdot 23,37 (\text{średnia zmiennej AGM}) + 0,102 \cdot 41,81 (\text{średnia zmiennej PS}) = 6,017.$$

Widzimy, że menedżerowie charakteryzują się wyższą przeciętną wartością różnicującą w porównaniu z pracownikami budżetowymi. Wykorzystując wyznaczone wartości, a dokładnie uśredniając je, możemy sformułować prostą regułę klasyfikacyjną, umożliwiającą przewidywanie przynależności jednostki do danej grupy: $\bar{Y} = (7,736 + 6,017)/2 = 6,877$. Jeżeli wartość jednostkowa jest wyższa od 6,877, osobę należy sklasyfikować jak menedżera, natomiast gdy jest od 6,877 niższa – jako pracownika sfery budżetowej.

Zaprezentowany w artykule przykład analizy wielozmiennowej wyników badań Strzałeckiego i Tomaszewicz nad profesjonalizmem menedżerskim jest przykładem realnie przeprowadzonych badań, w których konstrukt teoretyczny został opisany za pomocą 15 wymiarów i jako pojęcie nie może być zredukowany do mniejszej ich liczby na poziomie planowania następnych badań. Nawet teraz, gdy metodą *stepwise* wygenerowano zbiór zmiennych najsilniej dyskryminujących obydwie populacje, redukcja wymiarów konstrukt nie jest uzasadniona, zważywszy na wyniki MANOVA; nie tylko wyniki statystyki wielowymiarowej są istotne, ale aż czternaście z analizowanych zmiennych (traktowanych jednowymiarowo, bez rozważania ich struktury powiązań) ma istotnie różne wartości oczekiwane w obydwu populacjach (por. Tabela 3). Równocześnie aż siedem zmiennych w klasycznym rozwiązaniu analizy dyskryminacyjnej (por.

Tabela 4) ma wagi wyższe niż 0,3. Znalezione różnice nie upoważniają do odrzucenia tej definicji konstrukt na rzecz definicji bardziej redukcijnej, mimo że – jak wskazuje na to synteza wyników wszystkich przeprowadzonych analiz – osiową rolę w jego przypadku pełnią dwa ujawnione wymiary.

Zaprezentowany wyżej przykład ilustruje dwugrupowy plan badawczy obejmujący więcej niż jedną zmienną objaśnianą, będący podstawowym planem badawczym wymagającym zastosowania MANOVA i analizy dyskryminacyjnej. Dla schematów bardziej skomplikowanych łatwo wyprowadzić procedury obliczeniowe, uogólniając wzory podane wyżej.

Podsumowanie

W powyższym tekście – po szczegółowym opisanu na poziomie formalnym wielowymiarowej analizy wariancji i przedstawieniu wszelkich koniecznych obliczeń na fikcyjnych danych – odwołano się do danych rzeczywistych, uzyskanych z konkretnych badań psychologicznych. Chodziło o spełnienie dwu celów: po pierwsze o ilustrację przekształcania wartości uzyskanych statystyk we właściwe interpretacje psychologiczne oraz – po drugie – o przekonanie Czytelnika do konieczności używania modelu analizy dyskryminacyjnej po MANOVA jako metody wobec niej komplementarnej.

W pewnym sensie analiza dyskryminacyjna może być traktowana jako specyficzny wielozmiennowy test kontrastów niezbędnie stosowany po MANOVA, pełniący podobną rolę w przypadku interpretacji danych, jak jednoczynnowe testy kontrastów stosowane po ANOVA, czy analizach z powtarzanymi pomiarami.

Podczas opisu metody MANOVA starano się pokazać, na czym polega jej rozszerzenie względem analizy jednowymiarowej, czyli metody ANOVA, ze szczególnym zwróceniem uwagi na specyficzne zalety MANOVA. Przewaga wielowymiarowej analizy wariancji (MANOVA) nad jednowymiarową (ANOVA) – w kontekście obserwowania wielu zmiennych objętych siecią różnych relacji – jest oczywista z kilku powodów.

Po pierwsze, wyniki analiz jednowymiarowych nie mogą być syntetyzowane bezpośrednio (łączone w jedną wspólną konkluzję) z przyczyn statystycznych (teoretycznych), a mianowicie na skutek niemożności określenia wspólnej jednej wartości błędu wnioskowania pierwszego rodzaju. Jak wiadomo, w naukach empirycznych każdy wniosek jest probabilistyczny przy określonym błędzie wnioskowania. Wystarczy porównać przedstawione wyżej wzory dla stopni swobody statystyk F wielowymiarowych ze znanymi stopniami swobody dla ich jednowymiarowej wersji, by zdać sobie sprawę z rozbieżności rozkładów

prawdopodobieństwa F Fishera–Snedecora, którymi posługujemy się w obydwu sytuacjach (tym samym z rozbieżności wartości krytycznych, których przecież nie sposób uśredniać dla wielu decyzji).

Po drugie, wyniki analiz jednowymiarowych uzyskiwane (zawsze!) przy założeniu, że zmienne objaśniane są niezależne, są fałszywe tym bardziej, im bardziej powiązana jest analizowana struktura cech. Pokazał to opisany wyżej przykład. W wersji jednowymiarowej zmienna Poczucie zaradności miała wyższy wynik t -Studenta i d -Cohena niż zmienna Poczucie sensowności. Można byłoby przypuszczać, że ta pierwsza silniej jest dyskryminowana przez poziomy czynnik, albo – inaczej mówiąc – ta pierwsza wnosi większy wkład w rozbieżność między wynikami obydwu grup. Analizy wielowymiarowe, uwzględniające siłę powiązań między zmiennymi, obaliły tę konkluzję. Uwypukla to konieczność analizowania struktury zależności między zmiennymi objaśnianymi. W przypadku analiz jednowymiarowych interpretowanie różnic między wartościami średnich wymaga uwzględnienia ich wariancji, natomiast przeprowadzanie analiz wielowymiarowych bez odwołania się do zależności łączących objaśniane zmienne jest nie tylko bezcelowe, lecz wręcz bezsensowne.

Po trzecie, wielokrotnie stosowane – dla każdej zmiennej oddzielnie – jednowymiarowe i odpowiednio czynnikowe analizy wariancji (ANOVA) traktowane jako kontrasty parametrów pokażą tylko, jaki jest samodzielny wkład każdej zmiennej do różnicy między grupami. Natomiast stosowana zawsze po wielowymiarowej analizie wariancji analiza dyskryminacyjna odtworzy cząstkowy wkład każdej zmiennej z grupy łącznie analizowanych zmiennych związanych konkretną siecią wartości współczynników r -Pearsona, czyli konkretnie wzajemnie się determinujących.

Powyższe argumenty nie powinny stanowić zachęty do ignorowania rozwiązań jednowymiarowych. W interpretacji przykładu dobitnie wskazano, że wszystkie informacje są ważne i dopiero ujęte komplementarnie stanowią tło formalne dla decyzji merytorycznej.

LITERATURA CYTOWANA

- Aczel, A. D. (2000). *Statystyka w zarządzaniu. Pełny wykład*. Warszawa: Wydawnictwo Naukowe PWN.
- Ahrens, H., Läuter, J. (1979). *Wielowymiarowa analiza wariancji*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46.
- Aranowska, E. (1987). Elementy zastosowań modelu wielowymiarowej analizy wariancji (MANOVA) w badaniach psychologicznych. W: J. Brzeziński (red.), *Wielozmiennowe modele statystyczne w badaniach psychologicznych* (s. 113–151). Warszawa–Poznań: Państwowe Wydawnictwo Naukowe.
- Aranowska, E. (1996). *Metodologiczne problemy zastosowań modeli statystycznych w psychologii. Teoria i praktyka*. Warszawa: Studio 1.
- Aranowska, E. (2005). *Pomiar ilościowy w psychologii*. Warszawa: Wydawnictwo Naukowe Scholar.
- Aranowska, E., Ciok, A. (1992). Związek między zmiennymi w interpretacji analizy składowych głównych i analizy korespondencji. W: E. Aranowska (red.), *Wybrane problemy metodologii badań* (s. 133–168). Warszawa: Wydawnictwo Uniwersytetu Warszawskiego.
- Bathke, A. C., Harrar, S. W., Laurence, V., Madden, L. V. (2008). How to compare small multivariate samples using nonparametric tests. *Computational Statistics and Data Analysis*, 52, 4951–4965.
- Bergen, F. H., Seling, M. J. (1978). Uses of discriminant analysis following MANOVA: Multivariate statistics for multivariate purposes. *Journal of Applied Psychology*, 6, 689–697.
- Bollen, K., Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314.
- Brzeziński, J. (2006). Jaka psychologia? Jaka praktyka psychologiczna? Jakie kształcenie psychologów? W: E. Aranowska, M. Goszczyńska (red.), *Człowiek wobec wyzwań i dylematów współczesności* (s. 257–273). Warszawa: Wydawnictwo Naukowe Scholar.
- Brzeziński, J., Stachowski, R. (1984). *Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 6, 426–443.
- Cohen, J. (1977). *Statistical power analysis for behavioral sciences*. London: Academic Press.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Cole, D. A., Maxwell, S. E., Arvey, R., Salas, E. (1993a). Multivariate group comparisons of variable system: MANOVA and Structural Equation Modelling. *Psychological Bulletin*, 114, 174–184.
- Cole, D. A., Maxwell, S. E., Arvey, R., Salas, E. (1993b). How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin*, 115, 465–474.
- DasGupta, S., Perlman, M. D. (1973). On the power of Wilks' U-test for MANOVA. *Journal of Multivariate Analysis*, 3, 220–225.
- Doornik, J. A., Hansen, H. (2008). An omnibus for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, 70, 927–939.
- Enders, C. K. (2003). Performing multivariate group comparisons following a statistically significant MANOVA. *Measurement and Evaluation in Counseling and Development*, 36, 40–56.

- Everitt, B. S. (1979). A Monte Carlo investigation of the robustness of Hotelling's one and two sample T2 tests. *Journal of the American Statistical Association*, 74, 48–51.
- von Eye, A., Bogat, G. A. (2004). Testing the assumption of multivariate normality. *Psychology Science*, 46, 243–258.
- Feller, W. (1978). *Wstęp do rachunku prawdopodobieństwa* (t. 2). Warszawa: Państwowe Wydawnictwo Naukowe.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Finch, H. (2005). Comparison of the performance of non-parametric and parametric MANOVA test statistics when assumptions are violated. *Methodology*, 5, 27–38.
- Finch, H., Davenport, T. (2009). Performance of Monte Carlo permutation and approximate tests with multivariate means comparison with small sample sizes when parametric assumptions are violated. *Methodology*, 5, 60–70.
- Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations*. New York: John Wiley.
- Graham, J. M. (2008). The General Linear Model as Structural Equation Modeling. *Journal of Educational and Behavioral Statistics*, 33(4), 485–506.
- Hair, J. F., Anderson, R. E., Tatham, R. L., Black, W. C. (1995). *Multivariate data analysis with readings*. Englewood Cliffs, NJ: Prentice Hall.
- Harris, R. J. (1975). *A primer of multivariate statistics*. New York: Academic Press.
- Huberty, C. J. (1989). Problems with stepwise methods: Better alternatives. W: B. Thompson (red.), *Advances in social science methodology* (t. 1, s. 43–70). Greenwich, CT: JAI Press.
- Huberty, C. J. (1994). *Applied discriminant analysis*. New York: Wiley and Sons.
- Huberty, C. J., Smith, J. D. (1982). The study of effects in MANOVA. *Multivariate Behavioral Research*, 17, 417–432.
- Huberty, C. J., Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, 105, 302–308.
- Johnson, R. A., Wichern, D. W. (2002). *Applied multivariate statistics*. Upper Saddle River, NJ: Prentice Hall.
- Konarski, R. (2004). Analiza zmiany z zastosowaniem analizy latentnych krzywych rozwojowych. *ASK. Społeczeństwo. Badania. Metody*, 13, 87–120.
- Koronacki, J., Mielniczuk, J. (2001). *Statystyka dla studentów kierunków technicznych i przyrodniczych*. Warszawa: Wydawnictwo Naukowo-Techniczne.
- Liang, J., Tang, M. (2009). Generalized F-tests for multivariate normal mean. *Computational Statistics and Data Analysis*, 53, 1177–1190.
- Mardia, K. V. (1980). Tests of univariate and multivariate normality. W: P. R. Krishnaiah (red.), *Handbook of statistics* (t. 1, s. 279–320). Amsterdam: North Holland.
- McArdle, B. H., Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, 82, 290–297.
- Mielke, P. W., Berry, K. J. (2007). *Permutation methods: A distance function approach*. New York: Springer-Verlag.
- Morrison, D. F. (1990). *Wielowymiarowa analiza statystyczna*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of American Statistical Association*, 69, 894–908.
- Olson, C. L. (1976). On choosing a test statistics in multivariate analysis of variance. *Psychological Bulletin*, 4, 579–586.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., Briggs, N. E. (2008). *Latent Growth Curve Modeling*. Los Angeles–London: SAGE.
- Rao, C. R. (1994). *Statystyka i prawda*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Raykov, T. (2001). Testing multivariate covariance structure and means hypotheses via structural equation modeling. *Structural Equation Modeling*, 8, 224–256.
- Raykov, T., Marcoulides, G. A. (2008). *An introduction to applied multivariate statistics*. New York: Routledge.
- Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. W: B. Thompson (red.), *Advances in educational research: Substantive findings, methodological developments* (t. 1, s. 99–105). Greenwich, CT: JAI Press.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences*. New Jersey: Lawrence Earlbaum.
- Strzałeczki, A. (2003). *Psychologia twórczości. Między tradycją a ponowoczesnością*. Warszawa: Wydawnictwo UKSW.
- Strzałeczki, A., Tomaszewicz, B. (2004). Wyznaczniki efektywności funkcjonowania polskich menedżerów w warunkach zmian systemowych. *Studia Psychologica*, 5, 125–139.
- Tabachnik, B. G., Fidell, L. S. (2001). *Using multivariate statistics*. New York: HarperCollins.
- Tang, K. L., Algina, J. (1999). Performance of four multivariate tests under variance-covariance heteroscedascity. *Multivariate Behavioral Research*, 4, 391–405.
- Tatsuoka, M. M. (1971). *Multivariate analysis: Techniques of education and psychological research*. New York: John Wiley.
- Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology*. Monterey: Brookes/Cole Publishing Company.
- Todorov, V., Filzmoser, P. (2010). Robust statistics for the one-way MANOVA. *Computational Statistics and Data Analysis*, 54, 37–48.
- Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development*, 21(4), 146–148.
- Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement*, 55(4), 525–534.
- Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education*, 70(1), 80–93.
- Zieliński, R. (1985). *Metody odporne w statystyce*. Wykład na XII Konferencji Naukowo-Szkoleniowej „Statystyka Matematyczna”. Białejewko.

PRZYPISY

1. Wektor jest to odcinek w przestrzeni, mający określoną długość i kierunek. Dla zaznaczenia kierunku mówi się o początku i końcu wektora, oznaczając go strzałką. Za początek wektorów obserwacji uważa się początek układu współrzędnych. Zbiór punktów próbkowych to końce wektorów obserwacji. Długość wektora (długość odcinka) we współrzędnych prostokątnych równa jest pierwiastkowi z sumy kwadratów jego współrzędnych.

2. Główną przekątną jest ta z dwu przekątnych macierzy, którą wyznaczają elementy niezmiennicze położenia przy zamianie pierwszego wiersza z pierwszą kolumną, drugiego wiersza z drugą kolumną itd., tzn. elementy $\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp}$ macierzy (2). Macierz, która powstaje z macierzy A przez zamianę kolumn na

wiersze, nazywa się macierzą transponowaną (lub przestawioną) względem A i z reguły oznacza się ją symbolem A^T .

3. Kowariancja zmiennej samej z sobą równa jest wariancji tej zmiennej, toteż chociaż na głównej przekątnej macierzy Σ są wariancje, mówi się poprawnie: macierz kowariancji.

4. W algebrze macierzy nie istnieje operacja dzielenia macierzy, natomiast można pomnożyć jedną macierz przez macierz odwrotną do drugiej. Ponieważ oszacowanie mianownika stanowi macierz G , więc jest odwrotność G^{-1} występuje w teście \tilde{F} . Z pomnożenia tych dwu macierzy otrzymuje się także macierz. Skrót *tr* (od ang. słowa *trace* – ślad) oznacza sumę tych dwu elementów leżących na głównej przekątnej macierzy. Zatem do testu \tilde{F} w każdej z tych trzech sytuacji wstawia się nie macierz, lecz sumę elementów głównej przekątnej macierzy uzyskanej z wymnożenia odpowiednich H i G^{-1} .

Multivariate Analysis of Variance – MANOVA

Elżbieta Aranowska¹, Jolanta Rytel²

¹ Faculty of Psychology, University of Warsaw

² Institute of Psychology, Cardinal Stefan Wyszyński University in Warsaw

Abstract

The paper describes a model of the multivariate analysis of variance (MANOVA). First, we outline differences between this method and the univariate analysis of variance (ANOVA). We present basic repeated measures designs and point to the research designs that provide data which can be analyzed only with the MANOVA models. We describe formal structure of the MANOVA model and provide its basic definitions. We show how these definitions are related to terms of ANOVA. Development of the ANOVA logic into the MANOVA model is shown in relation to a discussion on independence between expected values of variables and their bivariate correlations (more precisely – means of variables and Pearson product-moment correlation coefficients). We explain how the assumptions, null hypotheses and test statistics of MANOVA have been developed from the ANOVA model. We point to the inconclusiveness of the formal MANOVA solution (lack of the one, established, test statistic) and show these test statistics which appeared most often in the statistical software in the last twenty years. We illustrate formalities of the model with one fictional example of a simple one-way MANOVA. All test statistics introduced in this paper were calculated by hand and compared with SPSS output. Moreover, an example of application of multivariate analysis of variance in psychological research was portrayed, using a study on evaluation of managers' performance. In this example, we emphasize reasons why it is necessary to complement MANOVA with another method: discriminant analysis.

Key words: Multivariate Analysis of Variance – MANOVA, discriminant analysis, multivariate methods, statistical models of data analysis

Złożono: 26.04.2010

Zaakceptowano do druku: 3.10.2010