PRZEGLĄD PSYCHOLOGICZNY, 2004, TOM 47, Nr 4, 367-378

Analiza kontrastów: między eksploracją a testowaniem hipotez[1]

Tytus Sosnowski[2]
Wydział Psychologii Uniwersytetu Warszawskiego

CONTRAST ANALYSIS: BETWEEN EXPLORATION AND TESTING HYPOTHESES

Abstract. The article discusses methods of comparison among means in experimental designs, where independent variable has more than two levels. These differences can be analyzed in two ways: as planned (*a priori*) contrasts or unplanned (*post hoc, post mortem*) contrasts. The choice of the method of analysis depends on the research problem. If the aim is to check whether differences among means are consistent with predictions, planned contrasts will be the proper choice. In such a case the omnibus F test does not provide any useful information on the tested hypotheses and can be omitted. Instead of ANOVA, one can use multiple regression analysis, which allows for direct assessment of previously defined contrasts. If, however, no precise predictions concerning differences among means exist, the proper method of analysis will be the omnibus F test followed (if F will turn out significant) by post hoc comparisons among means. The difference between the two kinds of contrasts corresponds to two fundamental research strategies: testing hypotheses and exploration.

Celem artykułu jest przegląd metod analizy różnic między średnimi grupowymi w planach eksperymentalnych, w których zmienna niezależna ma więcej niż dwa poziomy, a charakterystyka zmiennej zależnej upoważnia do stosowania statystyk parametrycznych, takich jak analiza wariancji czy analiza regresji wielokrotnej. Aby uczynić omawiane zagadnienia teoretyczne bardziej przystępnymi, odwołuję się do procedur obliczeniowych i statystyk dostępnych w *menu* pakietu statystycznego SPSS (wersja 10.1).

1 Praca finansowana z programu badawczego BST 843/29.
2 Adres do korespondencji: Wydział Psychologii, Uniwersytet Warszawski, ul. Stawki 5/7, 00-183 Warszawa; e-mail: tytus@engram.psych.uw.edu.pl

TYTUS SOSNOWSKI

KONTRASTY *POST HOC*

W przypadku eksperymentu dwugrupowego ocena istotności różnic między średnimi jest prosta: możemy użyć do tego celu test t Studenta lub test F Snedecora (przy dwóch grupach F równa się dokładnie t^2). Jeśli jednak czynnik eksperymentalny ma więcej niż dwa poziomy, porównywanie średnich staje się problemem dużo bardziej złożonym. Wybór odpowiedniej metody analizy zależy bowiem nie tylko od kryteriów czysto statystycznych, ale też od sposobu sformułowania pytania badawczego.

Wyobraźmy sobie fikcyjny eksperyment, w którym badacz analizuje wpływ zapachów na efektywność wykonywania zadań umysłowych. Cztery grupy eksperymentalne rozwiązują to samo zadanie, każda jednak w obecności innego zapachu. Załóżmy, że badacz nie sformułował żadnej hipotezy kierunkowej, lecz pragnie jedynie uzyskać odpowiedź na pytanie, czy zapachy wpływają na poziom wykonania zadania. Przy tak sformułowanym pytaniu właściwym postępowaniem będzie zastosowanie analizy wariancji i oszacowanie istotności zróżnicowania średnich za pomocą testu F. Jeśli wartość F okaże się nieistotna, nie będzie można odrzucić hipotezy zerowej mówiącej o braku różnic między oczekiwanymi wartościami średnich (Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$). W rezultacie nie będzie podstaw do stwierdzenia, że zapach wpływa na efektywność procesów umysłowych. Jeśli natomiast test F okaże się istotny, badacz będzie mógł wnioskować, że zapach wpływa na poziom wykonania zadania.

Istotność zbiorczego testu F (*omnibus F test*), czyli testu F zastosowanego do oceny różnicowania więcej niż dwóch średnich jednocześnie, oznacza, że przynajmniej jedna różnica między średnimi jest istotna statystycznie. Ponieważ nie sformułowano hipotezy, między którymi średnimi taka różnica ma zachodzić, logicznym krokiem będzie porównanie każdej średniej z każdą inną. Jeśli liczba średnich równa się k, liczba możliwych porównań między pojedynczymi średnimi grupowymi wynosi $k(k\text{-}1)/2$. Przy czterech grupach będziemy więc mieli 4(3)/2 = 6 możliwych porównań.

Wiadomo, że do wielokrotnych porównań między średnimi po teście F nie powinno się stosować testu t ani testu F, ponieważ grozi to zbyt pochopnym odrzucaniem hipotezy zerowej. Jeśli liczba grup, a tym samym liczba możliwych porównań wzrasta, to wzrasta też prawdopodobieństwo pojawienia się krytycznej wartości testu t lub F w sposób losowy. Aby uniknąć zbyt pochopnego odrzucenia hipotezy zerowej, do analizy różnic międzygrupowych po teście F skonstruowano specjalne testy, określane mianem kontrastów nieplanowanych (inaczej: *post hoc* lub *post mortem).*

Kontrast *post hoc* jest pewną miarą różnicy między dwiema średnimi. Kontrasty takie różnią się tym od standardowego testu t lub F, że uwzględniają liczbę porównywanych średnich. Jeśli liczba porównań wzrasta, wzrasta też wartość krytyczna testu *post hoc*. Testy *post hoc* opierają się na różnych założeniach, których omówienie przekracza ramy tego artykułu. Informacje na ten temat można znaleźć m.in. w pracach Haysa (1981), Kirka (1995), Winera (1971) oraz Winera, Browna i Michelsa (1991), a w literaturze polskojęzycznej – w pracach Brzezińskiego (2000) oraz Fergusona i Takane (1997). W przypadku niektórych testów (np. testu Dunneta, testu uczciwie istotnej różnicy Tukeya, testu Scheffego), ustalana jest jedna krytyczna wartość testu dla wszystkich porównań między średnimi. W innych przypadkach (test Studenta-Newmana-Keulsa, test Newmana-Keulsa, test Duncana) krytyczna wartość testu zależy od odległości między średnimi w szeregu średnich uporządkowanych według ich wielkości: wartość ta jest niższa wówczas, gdy dwie średnie sąsiadują ze sobą, niż wtedy, gdy rozdzielają je inne średnie. Testy *post hoc* różnią się od siebie stopniem konserwatyzmu. Z dwóch testów bardziej konserwatywny jest ten, na podstawie którego trudniej odrzuca się hipotezę zerową. Najbardziej konserwatywnym testem *post hoc* jest test Scheffego – może on nie wykazać istotności różnic między średnimi, mimo że inny test *post hoc* będzie w tych samych warunkach istotny. Z drugiej strony, jeśli test Scheffego okaże się istotny, każdy inny test będzie również istotny. Pakiet statystyczny SPSS oferuje kilkanaście kontrastów *post hoc*, w tym wszystkie wymienione powyżej.

Metodą analizy kontrastów nieplanowanych, o której warto dodatkowo wspomnieć, jest procedura Bonferroniego (por. Hays, 1981; Pedhazur, 1982; Winer, Brown, Michels, 1991). Polega ona na zastosowaniu zwykłego testu t lub testu F, ale z pewną poprawką: poziom istotności dla standardowego testu t lub F mnożony jest przez liczbę dokonywanych porównań między średnimi (warto podkreślić, że bierzemy tu pod uwagę liczbę porównań faktycznie dokonywanych, a nie wszystkich możliwych). Załóżmy, że mamy trzy średnie grupowe: , i , i dokonujemy za pomocą testu t porównania między średnimi i oraz między średnimi i . Jeśli prawdopodobieństwo błędu α dla standardowego testu t użytego do oceny obu różnic byłoby równe 0,05 i 0,001, to po uwzględnieniu poprawki Bonferroniego wynosiłoby ono 0,05 x 2 = 0,1 oraz 0,001 x 2 = 0,002. Procedura Bonferroniego jest prosta w użyciu. Ponadto można z niej korzystać zarówno przy porównaniach średnich grupowych, jak i przy porównaniach średnich dla powtarzanych pomiarów (korygując poziom istotności testu t dla danych skorelowanych). Z drugiej strony, procedura Bonferroniego jest bardzo konserwatywna, zwłaszcza gdy liczba porównań jest duża. Jej stosowanie może więc prowadzić do zbyt pochopnej decyzji o nieodrzuceniu hipotezy zerowej, gdy jest ona faktycznie fałszywa.

ANALIZA KONTRASTÓW: MIĘDZY EKSPLORACJĄ A TESTOWANIEM HIPOTEZ

KONTRASTY PLANOWANE

Wyobraźmy sobie inny fikcyjny eksperyment. Badacz sformułował hipotezę teoretyczną, zgodnie z którą pewien związek chemiczny, występujący w niektórych substancjach zapachowych, działa jako stymulant i zwiększa efektywność procesów korowych. Badacz chce sprawdzić zasadność swej hipotezy. W czterogrupowym eksperymencie zastosował cztery bodźce zapachowe, z których dwa (A i B) były stymulantami, natomiast dwa inne (C i D) nie miały takich właściwości.

Jeśli celem eksperymentu jest sprawdzenie, czy dane empiryczne potwierdzają sformułowaną wyżej hipotezę teoretyczną, to odpowiedzi na to pytanie nie udzieli zbiorczy test *F*. Istotność *F* wskazałaby co prawda na istnienie różnic między średnimi, nie wiedzielibyśmy jednak, czy różnice te są zgodne, czy też niezgodne z przewidywaniami wynikającymi z teorii. Ponieważ teoria przewiduje wystąpienie dokładnie określonych różnic, nie ma też sensu porównywanie każdej średniej z każdą inną. Właściwym postępowaniem będzie w tym wypadku analiza wybranych kontrastów planowanych (inaczej – kontrastów *a priori*). Dla przetestowania hipotezy teoretycznej, będącej podstawą opisanego wyżej eksperymentu, moglibyśmy np. zdefiniować trzy kontrasty:

(1) Porównanie średnich A i B ze średnimi C i D, które udzieliłoby odpowiedzi na pytanie, czy wykonywanie zadań umysłowych w obecności stymulantów daje lepsze wyniki niż wykonywanie ich w obecności zapachów obojętnych.

(2) Porównanie średniej A ze średnią B – dla sprawdzenia, czy nie ma różnic między oddziaływaniem obu stymulantów.

(3) Porównanie średniej C ze średnią D – dla sprawdzenia, czy nie ma różnic między oddziaływaniem obu bodźców neutralnych.

Analiza kontrastów planowanych nie nastręcza większych problemów, jeśli możemy posłużyć się pakietem statystycznym, takim jak SPSS for Windows. Użytkownik SPSS może sam zdefiniować potrzebne mu kontrasty (opcja: „porównanie średnich > jednoczynnikowa analiza wariancji”) lub wybrać z menu programu jeden z gotowych zestawów kontrastów (opcja: „ogólny model liniowy”). Samodzielne definiowanie kontrastów polega na przypisaniu średnim grupowym odpowiednich współczynników, które mogą przybierać dowolne wartości, całkowite lub ułamkowe, choć najczęściej, dla wygody, wybieramy liczby całkowite i możliwie najmniejsze. Zazwyczaj żąda się też (i taką dyrektywę znajdziemy również w SPSS), aby suma współczynników definiujących dany kontrast była równa zero (na temat konsekwencji niespełnienia tego warunku zob. Hays, 1981). Oznacza to, że niektóre współczynniki muszą mieć wartość dodatnią, inne zaś – ujemną. Średnie, którym przypisaliśmy współczynniki dodatnie, będą porównywane ze średnimi, którym przypisaliśmy współczynniki ujemne. Jeśli natomiast przypiszemy jakiejś średniej współczynnik 0, będzie ona pominięta w analizie. Przykładowo, zestaw współczynników: 1, -1, 0, 0 definiuje kontrast jako różnicę między średnią dla grupy pierwszej i średnią dla grupy drugiej, z pominięciem grup trzeciej i czwartej. Zestaw współczynników: 3, -1, -1, -1 określa natomiast różnicę między pierwszą grupą a trzema pozostałymi. Chcąc zdefiniować kontrasty w naszym eksperymencie (te, które przedstawiliśmy wyżej w punktach 1, 2 i 3), moglibyśmy użyć następujących współczynników: (1) 1, 1, -1, -1; (2) 1, -1, 0, 0; i (3) 0, 0, 1, -1.

Wielkość *i*-tego kontrastu jest ważoną sumą średnich, tj. sumą analizowanych średnich, pomnożonych przez przypisane im współczynniki kontrastu: . Przykładowo, wielkość kontrastu zdefiniowanego za pomocą współczynników: 1, -1, 0, 0 będzie równa: ψ_i = (1) + (-1) + (0) + (0).

Istotność kontrastu może być testowana za pomocą statystyki *F* lub statystyki *t*. Jeśli mamy do czynienia z planem w grupach zrandomizowanych, grupy są równoliczne, a ich wariancje homogeniczne, jako mianownika testu *F* możemy użyć średniego kwadratu wewnątrz grup – tego samego, jakiego używamy w analizie wariancji przy szacowaniu istotności zróżnicowania między grupami. Licznikiem testu *F* będzie średni kwadrat dla kontrastu, czyli suma kwadratów dla kontrastu (sposób jej obliczania podajemy dalej), podzielona przez 1 stopień swobody. Test *t* z kolei jest stosunkiem wielkości danego kontrastu ψ_i do jego błędu standardowego $S_{\psi i}$ – wzory na obliczenie $S_{\psi i}$ można znaleźć w pracach Kirka (1995) oraz Winera, Browna i Michelsa (1991). Ponieważ *F* dla kontrastu ma zawsze jeden stopień swobody w liczniku, możemy go łatwo przekształcić w *t*: $F = t^2$.

W *menu* pakietu SPSS (opcja: „ogólny model liniowy”) dostępnych jest sześć gotowych zestawów kontrastów planowanych. Kontrasty odchylenia (*deviation*) polegają na porównaniu każdej grupy (z wyjątkiem grupy odniesienia) ze średnią całkowitą. Kontrasty proste (*simple*) umożliwiają porównanie jednej wybranej grupy (może to być np. grupa kontrolna) z każdą z pozostałych grup. Kontrasty różnicy (*difference*) polegają na porównaniu każdej grupy (z wyjątkiem pierwszej) z grupą poprzednią, natomiast kontrasty powtórzone (*repeated*) – na porównaniu każdej grupy (z wyjątkiem ostatniej) z grupą następną. Kontrast Helmerta to porównanie średniej każdej grupy (z wyjątkiem grupy ostatniej) ze średnią ze wszystkich następujących po niej grup. Kontrasty wielomianowe (*polynomial*) pozwalają ocenić, czy średnie grupowe dadzą się dobrze przybliżyć za pomocą linii będącej wykresem wielomianu określonego stopnia. Kontrasty te zostaną omówione dokładniej nieco później. Jeśli chcemy policzyć inne kontrasty, musimy je zdefiniować samodzielnie, przypisując średnim grupowym odpowiednie współczynniki. Samodzielne definiowanie kontrastów w SPSS jest możliwe po wybraniu opcji: „porównanie średnich > jednoczynnikowa analiza wariancji".

Planowane kontrasty ortogonalne

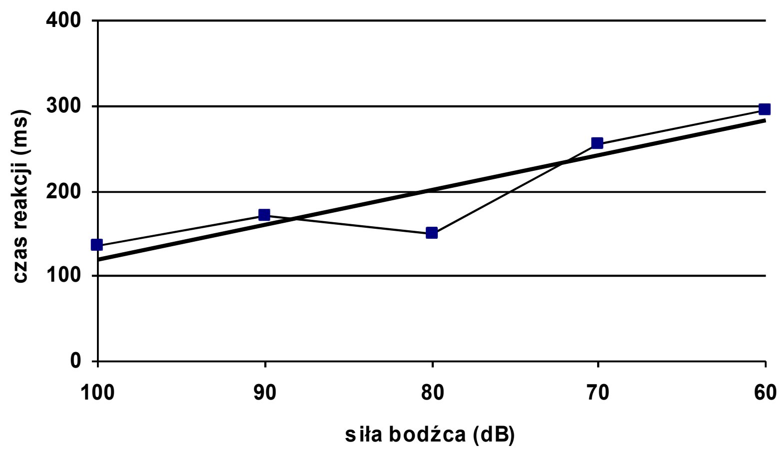
Kontrasty planowane można podzielić na ortogonalne i nieortogonalne. Kontrasty są ortogonalne, jeśli dostarczają niezależnych informacji o różnicach między średnimi. Jeżeli grupy są równoliczne, dwa kontrasty są ortogonalne względem siebie wtedy, gdy iloczyny ich współczynników, przypisanych tym samym grupom, sumują się do zera. Jeśli mamy więcej niż dwa kontrasty, są one wzajemnie ortogonalne, gdy wszystkie możliwe pary kontrastów są ortogonalne. Przeanalizujmy trzy przykładowe kontrasty:
(1) 1, -1, 0, 0;
(2) 1, 0, -1, 1;
(3) 1, 0, 0, -1.
Suma iloczynów dla kontrastów 1 i 2 wynosi: (1)(1) + (-1)(0) + (0)(-1) + (0)(1) = 1. Ponieważ suma ta jest różna od zera, kontrasty 1 i 2 nie są ortogonalne względem siebie. Tym samym i cały zestaw trzech kontrastów nie jest ortogonalny. Przeanalizujmy teraz inny zestaw kontrastów:
(1) +3, -1, -1, -1;
(2) 0, +2, -1, -1;
(3) 0, 0, +1, -1.
Suma iloczynów współczynników dla kontrastów 1 i 2 wynosi:
(+3)(0) + (-1)(+2) + (-1)(-1) + (-1)(-1) = 0; dla kontrastów 1 i 3 wynosi:
(+3)(0) + (-1)(0) + (-1)(+1) + (-1)(-1) = 0; dla kontrastów 2 i 3 wynosi:
(0)(0) + (+2)(0) + (-1)(+1) + (-1)(-1) = 0.
Ponieważ wszystkie sumy są równe zeru, trzy podane wyżej kontrasty są względem siebie ortogonalne.
Przy k grupach liczba kontrastów ortogonalnych nie może być większa od k-1. Jeśli kontrasty są wzajemnie ortogonalne, suma kwadratów związana z efektem zmiennej niezależnej równa się sumie sum kwadratów związanych z k-1 kontrastami. Jest to przypadek analogiczny do tego, z jakim mamy do czynienia w wielokrotnej analizie regresji, kiedy zmienne niezależne (w języku analizy regresji – predyktory) włączone do równania regresji są nieskorelowane ze sobą: suma kwadratów zmiennej zależnej, przewidywana na podstawie wszystkich zmiennych niezależnych łącznie, jest równa sumie sum kwadratów przewidywanych na podstawie poszczególnych zmiennych niezależnych. SPSS nie podaje oszacowania sum kwadratów dla poszczególnych kontrastów. Podaje jedynie wielkość kontrastów oraz ich istotność statystyczną. Znając wielkość kontrastu, współczynniki kontrastu oraz liczebność grup, można jednak łatwo obliczyć sumę kwadratów dla kontrastu. Jeśli grupy są równoliczne, wzór na sumę kwadratów jest dość prosty (zob. Ferguson, Takane, 1997, s. 345-346):
.

Symbol $SS_{\psi i}$ oznacza sumę kwadratów dla i-tego kontrastu, n – liczbę osób w grupie, ψ_i – wielkość i-tego kontrastu, ΣC_{ij}^2 – sumę kwadratów współczynników i-tego kontrastów dla j grup. Wzory obliczeniowe dla innych przypadków (w tym m.in. dla grup nierównolicznych) można znaleźć w podręczniku Winera, Browna i Michelsa (1991).
Wśród specjalistów nie ma zgody co do korzyści wynikających ze stosowania kontrastów ortogonalnych w porównaniu z kontrastami nieortogonalnymi (por. Hays, 1981; Rosenthal, Rosnow, Rubin, 2000). Zaletą kontrastów ortogonalnych jest to, że sumy kwadratów związane z poszczególnymi kontrastami można przedstawić w postaci procentu międzygrupowej sumy kwadratów. Uzyskujemy w ten sposób swoistą miarę wielkości poszczególnych kontrastów. Przypuśćmy, że badamy wpływ dwóch rodzajów wzmocnienia na szybkość uczenia się szczurów w labiryncie. Dla grupy 1 wzmocnieniem jest pokarm, dla grupy 2 – wyłączenie przykrego dźwięku, natomiast grupa 3 (kontrolna) nie otrzymuje żadnego wzmocnienia. Zdefiniowaliśmy dwa kontrasty ortogonalne: (1) 1, 1, -2 oraz (2) 1, -1, 0. Powiedzmy, że międzygrupowa suma kwadratów wynosi 300, z czego na pierwszy kontrast przypada 240, a na drugi 60. Na podstawie takich wyników moglibyśmy powiedzieć, że 80% zróżnicowania między grupami tłumaczy samo działanie wzmocnienia, natomiast różnica między obu rodzajami wzmocnienia – tylko 20% tego zróżnicowania.
Z drugiej strony, są sytuacje, kiedy uzasadnione jest stosowanie kontrastów nieortogonalnych. Ilustruje to dobrze przykład zaczerpnięty z pracy Rosenthala, Rosnowa i Rubina (2000, s. 156 n.). Przypuśćmy, że interesuje nas poziom wykonania pewnego zadania umysłowego przez dzieci w różnym wieku. Badamy cztery grupy dzieci, z których każda kolejna jest starsza od poprzedniej. Chcemy sprawdzić cztery alternatywne hipotezy. Zgodnie z pierwszą (kontrast: -3, -1, 1, 3) poziom wykonania będzie wzrastał liniowo z wiekiem; zgodnie z drugą (-1, -1,

-1, 3) każda z trzech grup młodszych uzyska wynik gorszy niż grupa najstarsza; zgodnie z trzecią (-1, 0, 0, 1) różnica wystąpi tylko między grupami skrajnymi (najmłodszą i najstarszą); zgodnie z czwartą natomiast (-1, -1, 1, 1) nie będzie różnicy między dwiema najmłodszymi ani między dwiema najstarszymi grupami, obie grupy najstarsze powinny zaś uzyskać wyniki lepsze od obu grup najmłodszych. Oszacowanie istotności wyżej wymienionych kontrastów pozwoliłoby nam ustalić, które ze sformułowanych wcześniej przewidywań potwierdziły się. Na tej podstawie moglibyśmy ocenić zasadność hipotez będących podstawą poszczególnych przewidywań. Nie moglibyśmy jednak porównywać względnej wielkości poszczególnych kontrastów, ponieważ nie są one ortogonalne – każdy z kontrastów wyjaśnia jakąś część międzygrupowej sumy kwadratów, która jest też wyjaśniana przez inny kontrast.

Planowane kontrasty wielomianowe

Szczególną formą kontrastów ortogonalnych są kontrasty wielomianowe (*polynomial*), zwane też trendami wielomianowymi. Analiza takich kontrastów ma na celu ocenę, czy średnie grupowe dadzą się przybliżyć za pomocą linii będącej wykresem wielomianu *n*-tego stopnia (ściśle rzecz biorąc – sprawdzamy, czy odchylenie średnich grupowych od wybranej linii trendu jest istotne statystycznie). Wykresem wielomianu stopnia pierwszego jest linia prosta (trend liniowy), wielomianu stopnia drugiego – linia z jednym zgięciem (trend kwadratowy), wielomianu stopnia trzeciego – linia zagięta dwa razy (trend sześcienny, inaczej kubiczny). Analiza trendów stopnia wyższego niż trzeci zdarza się w psychologii rzadko, trudno bowiem znaleźć teorię lub prawo empiryczne mówiące o tak skomplikowanej zależności między zmiennymi.



Rys 1. Analiza kontrastów

Zaletą trendów wielomianowych jest to, że pozwalają ująć zależność między zmienną niezależną i zależną w postaci jednej funkcji. Rysunek 1 przedstawia fikcyjne dane z pięciogrupowego eksperymentu badającego wpływ siły bodźca dźwiękowego na czas reakcji motorycznej. Gdybyśmy chcieli porównywać średnie grupowe parami, musielibyśmy dokonać co najmniej czterech takich porównań: , , i , a wszystkich możliwych porównań między pojedynczymi średnimi można by przeprowadzić 10: 5x4/2. Stosując analizę trendów, uzyskamy jeden syntetyczny wynik (w postaci testu *t* lub testu *F*), informujący o tym, czy wybrana linia (np. linia prosta w podanym wyżej przykładzie) stanowi dobre przybliżenie średnich grupowych. Gdyby okazało się, że trend liniowy stanowi rzeczywiście najlepsze przybliżenie średnich przedstawionych na rycinie, moglibyśmy wnioskować, że wzrost siły bodźca (przynajmniej w analizowanym przedziale jego wartości) powoduje liniowe skrócenie czasu reakcji. Odchylenie poszczególnych średnich od linii trendu (gdy jest on istotny) możemy natomiast traktować jako efekt działania czynników losowych.
Analiza trendów jest szczególnie warta polecenia wtedy, gdy liczba średnich jest duża. Przy małej liczbie średnich wnioskowanie o istnieniu trendu może być ryzykowne. Przy dwóch średnich, jeśli tylko różnią się one istotnie między sobą, zawsze otrzymamy trend liniowy. Gdybyśmy jednak uwzględnili większą liczbę poziomów zmiennej niezależnej, mogłoby się np. okazać, że zależność między zmiennymi jest krzywoliniowa. Należy też pamiętać, że analiza trendów opiera się na założeniu, że zmienna niezależna jest ciągła, a jej dyskretne

wartości, przypisane kolejnym grupom lub warunkom eksperymentalnym, rosną lub maleją o wartość stałą. Używając SPPS możemy, podobnie jak w przypadku innych kontrastów planowanych, skorzystać z gotowych kontrastów wielomianowych lub też zdefiniować je samodzielnie, przypisując średnim grupowym odpowiednie współczynniki. Przykładowo, przy czterech średnich zestaw współczynników -3, -1, 1, 3 definiuje trend liniowy, zestaw 1, -1, -1, 1 – trend kwadratowy, a zestaw -1, 3, -3, 1 – trend sześcienny. Zestawy współczynników potrzebnych do zdefiniowania trendów wielomianowych różnego stopnia i dla różnej liczby średnich podają w swym podręczniku Ferguson i Takane (1997, tabela J).

Trendy wielomianowe mogą być również analizowane jako kontrasty *post hoc*. Różnica między kontrastami wielomianowymi *a priori* i *post hoc* polega na innym oszacowaniu wariancji błędu. Ponieważ SPSS nie oferuje kontrastów wielomianowych *post hoc,* nie będę ich tu bliżej omawiać. Informacje na ten temat można znaleźć w pracach Haysa (1981) oraz Winera, Browna i Michelsa (1991).

ANALIZA KONTRASTÓW A WIELOKROTNA ANALIZA REGRESJI

Do analizy kontrastów planowanych w badaniach eksperymentalnych możemy użyć nie tylko analizy wariancji, ale również analizy regresji wielokrotnej. Przynależność badanych do poszczególnych grup eksperymentalnych musimy wówczas zakodować za pomocą specjalnych wektorów (zob. Brzeziński, 1996; Cohen, Cohen, 1975; Ferguson, Takane, 1997; Pedhazur, 1982; Pedhazur, Pedhazur Schmelkin, 1991). Dla zakodowania czynnika eksperymentalnego o k poziomach potrzebujemy zawsze, niezależnie od metody kodowania, k-1 wektorów. Każdy z wektorów definiuje pewien kontrast, czyli różnicę między pojedynczymi średnimi lub grupami średnich. W równaniu regresji każdy wektor występuje jako odrębna zmienna niezależna.

Jeśli nie interesuje nas całkowity efekt zmiennej niezależnej, ale ściśle określone kontrasty, analiza regresji daje rozwiązanie prostsze i bardziej eleganckie niż analiza wariancji. Wszystkie dane potrzebne do oceny kontrastów znajdziemy w tabeli przedstawiającej współczynniki równania regresji. Miarą wielkości danego kontrastu jest współczynnik b równania regresji, związany z odpowiednim wektorem (w wypadku niektórych metod kodowania, b jest równe liczbowo analizowanej różnicy między średnimi). Istotność współczynnika b (w SPSS oceniana za pomocą testu t) jest równoznaczna z istotnością odpowiadającego mu kontrastu. Zastosowanie analizy regresji do analizy danych eksperymentalnych, w tym zwłaszcza do analizy kontrastów planowanych, omawiam w innym artykule (Sosnowski, 2004).

UWAGI KOŃCOWE

Jak starałem się wykazać wyżej, wybór metody testowania różnic między średnimi zależy w dużej mierze od sposobu sformułowania pytania badawczego. Istotne znaczenie ma tu zwłaszcza rozróżnienie między podejściem eksploracyjnym a testowaniem hipotez.

Zgodnie z modelem hipotetyczno-dedukcyjnym (por. Hempel, 1966/2000) testowanie hipotez teoretycznych polega na dedukowaniu z nich przewidywań i poddawaniu ich testom empirycznym. W idealnej sytuacji teoria powinna umożliwiać przewidywania na tyle jednoznaczne, abyśmy mogli wszystkie wyniki, możliwe do uzyskania w danym badaniu, podzielić na dwa podzbiory: (1) wyniki zgodne z oczekiwaniami wynikającymi z teorii; (2) wyniki niezgodne z takimi oczekiwaniami. Wiedząc dokładnie, jakie wyniki przewiduje teoria, możemy je przedstawić w formie kontrastów planowanych. Analiza kontrastów planowanych ma w takim przypadku przewagę nad analizą kontrastów *post hoc*, i to co najmniej z dwóch powodów. Po pierwsze, prawdopodobieństwo odrzucenia H_0 jest większe wówczas, gdy analizujemy daną różnicę między średnimi jako kontrast planowany, niż wtedy, gdy analizujemy ją jako kontrast *post hoc*. Po drugie, analiza kontrastów planowanych jest podejściem bardziej eleganckim. Jeśli przewidujemy wystąpienie określonej różnicy lub określonych różnic między średnimi, to nie ma powodu, aby porównywać każdą średnią z każdą inną. Jeśli badacz dokonuje takich porównań, to nasuwa się wątpliwość, czy badanie ma rzeczywiście na celu testowanie sformułowanych wcześniej hipotez, czy też jest jedynie eksploracją problemu.

Wiele badań psychologicznych, w tym również badań eksperymentalnych, ma jednak charakter eksploracyjny. Są one nastawione nie na testowanie hipotez teoretycznych, ale na poszukiwanie nowych zależności empirycznych. Co prawda, w każdym badaniu można dopatrzeć się jakieś hipotezy. Jeśli badacz decyduje się poddać badaniu relację między wybranymi zmiennymi, to zazwyczaj uważa, że istnieją ku temu pewne podstawy teoretyczne. Mogą być one jednak na tyle słabe, że nie pozwalają na formułowanie dokładnych przewidywań. Nie wiadomo więc dokładnie, jaki wynik badania będzie zgodny z teorią, a jaki z nią niezgodny. Może być też tak, że badacz formułuje dokładne przewidywania (opierając się np. na wynikach innych podobnych badań), ale nie potrafi powiązać ich jednoznacznie z teorią. Uzyskanie wyniku niezgodnego z przewidywaniami nie musi więc również prowadzić do zakwestionowania teorii. Jeśli planując badania, badacz odwołuje się do teorii lub formułuje przewidywania w taki sposób, że nie wiadomo, jakie wyniki będą zgodne z teorią, a jakie z nią niezgodne, rezultatem badania może być tylko ustalenie pewnych zależności empirycznych. Postępowanie takie nazwałbym: „niejawną eksploracją" lub „badaniem eksploracyjnym o pozorach testowania hipotez".

Jeśli nie jesteśmy pewni, jakich różnic między średnimi powinniśmy oczekiwać, właściwym rozwiązaniem będzie obliczenie zbiorczego testu F, a następnie – jeśli okaże się on istotny, a zmienna niezależna ma więcej niż dwa poziomy – porównanie każdej średniej z każdą inną za pomocą kontrastów *ex post*. Jeżeli analiza taka doprowadzi do odkrycia nieprzewidywanych wcześniej zależności empirycznych (czyli istotnych statystycznie kontrastów *post hoc*), badacz powinien znaleźć dla nich jakąś interpretację. Może jej poszukiwać na gruncie istniejących już teorii lub podjąć próbę zbudowania odpowiedniej teorii samodzielnie. W przeciwieństwie do testowania hipotez teoretycznych, proces ich generowania traktowany był przez długi czas w metodologii nauk marginesowo i dopiero ostatnio sytuacja zaczęła ulegać pewnym zmianom. W szczególności wskazuje się (por. Brzeziński, 1978; Paszkiewicz, 1977; Kosnarewicz, 1989), że zaproponowane przez Reichenbacha rozróżnienie między kontekstem odkrycia i kontekstem uzasadnienia jest zbyt schematyczne i nie oddaje dobrze natury procesu badawczego, a metody analizy danych mogą wspomagać proces generowania hipotez teoretycznych. Jakkolwiek byśmy jednak patrzyli na ten proces, wynik empiryczny nie może być potwierdzeniem interpretacji teoretycznej sformułowanej *ex post*. Interpretacja taka może być traktowana wyłącznie jako hipoteza, która musi być poddana niezależnym testom. Jeśli badacz nie przeprowadza takich testów, to można podejrzewać, że nie przywiązuje znaczenia do wykrytej zależności empirycznej lub też nie traktuje poważnie zaproponowanej przez siebie interpretacji.

Wydaje się, że tym, co najlepiej różnicuje podejście eksploracyjne od testowania hipotez, jest stosunek badacza do wyników, które są istotne statystycznie, ale niezgodne z przewidywaniami. Jeśli badania mają na celu testowanie teorii, wynik niezgodny z przewidywaniami jest wynikiem negatywnym. Przyczyny odpowiedzialne za taki wynik mogą być oczywiście różne. Mogą nimi być zwłaszcza różne błędy metodologiczne, np. niewłaściwa operacjonalizacja zmiennych teoretycznych czy niedostateczna kontrola zmiennych ubocznych i zakłócających. Jeśli jednak nie da się znaleźć takich błędów, przyczyn negatywnego wyniku należy upatrywać w samej teorii. Badacz musi ją wówczas albo zmodyfikować, albo odrzucić. Logika i praktyka testowania teorii (zwłaszcza teorii o podstawowym znaczeniu dla danej dyscypliny naukowej) na podstawie danych empirycznych to zagadnienia niezwykle złożone (por. Carnap, 1966/2000; Hempel, 1966/2000; Kuhn, 2001; Lakatos, 1995; Popper, 2002). Nie ulega jednak wątpliwości, że próby rewizji teorii, wymuszone negatywnymi wynikami testów empirycznych, to jeden z podstawowych czynników postępu w nauce.

Jeżeli badanie ma charakter eksploracyjny, wynik istotny statystycznie – choć niezgodny z oczekiwaniami – może być sukcesem, gdyż może oznaczać, że badaczowi udało się odkryć jakąś zależność empiryczną. Dokonań takich nie należy w żadnym wypadku lekceważyć. Warto też pamiętać, że wiele wybitnych odkryć naukowych dokonanych zostało w sposób nieplanowany. Dobrym przykładem może tu być przygoda naukowa naszej wielkiej rodaczki Marii Skłodowskiej-Curie, która szukając w smółce uranowej nowego pierwiastka promieniotwórczego, odkryła w niej, zupełnie nieoczekiwanie, dwa takie pierwiastki, nazwane później polonem i radem (Curie, 1979). Aby fakt empiryczny stał się odkryciem naukowym, musi zostać jednak powiązany z teorią, ta zaś musi być poddana niezależnym testom empirycznym. Samo gromadzenie faktów, choćby najbardziej spektakularnych, nie poszerza jeszcze naszego rozumienia świata.

BILBIOGRAFIA

Brzeziński, J. (1978). *Metodologiczne i psychologiczne wyznaczniki procesu badawczego w psychologii*. Poznań: Wydawnictwo Naukowe UAM.

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

Brzeziński, J. (2000). *Badania eksperymentalne w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe Scholar

Carnap, R. (1966/2000). *Wprowadzenie do filozofii nauki*. Warszawa: Fundacja Aletheia.

Cohen, J., Cohen, B. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ.: Lawrence Erlbaum Associates, Publishers.

Curie, E. (1979). *Maria Curie*. Warszawa: PWN.

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

Hays, W. L. (1981). *Statistics* (wyd. 3). New York: Holt, Rinehart & Winston.

Hempel, C. G. (1966/2000). *Filozofia nauk przyrodniczych*. Warszawa: Fundacja Aletheia.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (wyd. 3). Pacific Grove: Brooks/Cole Publishing Company.

Kosnarewicz, A. (1989). Kilka uwag o dystynkcji kontekst odkrycia-kontekst uzasadnienia. [W:] J. Brzeziński, K. Łastowski (red.), *Filozoficzne i metodologiczne podstawy teorii naukowych* (s. 275-285). Warszawa–Poznań: PWN. Poznańskie Studia z Filozofii Nauki, t. 11.

Kuhn, T. S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
Lakatos, I. (1995). *Pisma z filozofii nauk empirycznych*. Warszawa: PWN.
Paszkiewicz, E. (1977). The context of discovery and the context of justification – an opposition or a complementarity? *Poznań Studies in the Philosophy of the Sciences and the Humanities*, 3, 256-264.
Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (wyd. 2). New York: Holt, Rinehart and Winston.
Pedhazur, E. J., Pedhazur Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale: Lawrence Erlbaum Associates, Publishers.
Popper, K. R. (2002). *Logika odkrycia naukowego*. Warszawa: Wydawnictwo Naukowe PWN.
Rosenthal, R., Rosnow, R. L., Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach*. Cambridge: Cambridge University Press.
Sosnowski, T. (2004). *Zastosowanie analizy wielokrotnej regresji liniowej do analizy danych eksperymentalnych*. *Psychologia-Etologia-Genetyka*, 9, 53-80.
Winer, B. J. (1971). *Statistical principles in experimental design* (wyd. 2). Tokyo: McGraw-Hill Kogakusha, Ltd.
Winer, B. J., Brown, D. R., Michels, K. M. (1991). *Statistical principles in experimental design* (wyd. 3). Boston: McGraw Hill.