

Jerzy Brzeziński

Eksperymentalne badanie zmian - metodologiczne problemy stosowania pretestu zmiennej zależnej

Przegląd Socjologiczny Sociological Review 37, 113-139

1989

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

JERZY BRZEZIŃSKI – POZNAŃ

EKSPERYMENTALNE BADANIE ZMIAN – METODOLOGICZNE PROBLEMY STOSOWANIA PRETESTU ZMIENNEJ ZALEŻNEJ

WPROWADZENIE

DWA WARIANTY

Przeprowadzone w niniejszej pracy analizy metodologiczne dotyczyć będą takich sytuacji badawczych, które można ująć w następujący schemat. Wedle pierwszego wariantu, badacz rejestruje jakiś stan początkowy interesującego go zjawiska (np. poziom lęku grupy pacjentów depresyjnych), następnie próbuje oddziaływać na to zjawisko (np. za pomocą psychoterapii ukierunkowanej na redukcję poziomu lęku). Badanie kończy rejestracja stanu końcowego (np. pomiar poziomu lęku pacjentów po przeprowadzeniu psychoterapii). Z kolei w drugim wariantcie badacz dzieli (losowo!) grupę osób – reprezentatywną dla danej populacji – na dwie podgrupy. Po dokonaniu pomiaru „wyjściowego” natężenia zmiennej zależnej (np. postawy wobec jakiegoś obiektu) badacz w sposób zróżnicowany traktuje osoby z obu grup. Jedna grupa poddana jest specjalnym oddziaływaniom (np. stosuje się wobec niej techniki indoktrynacji). Druga grupa poddana jest oddziaływaniom neutralnym względem interesującej badacza postawy. Badanie kończy ponowny pomiar natężenia zmiennej zależnej.

PORÓWNANIE STANU KOŃCOWEGO ZE STANEM POCZĄTKOWYM

Badania wyżej opisane, nastawione są na znalezienie odpowiedzi na pytanie o istotność różnicy między pierwszym i drugim pomiarem natężenia zmiennej zależnej. Chodzi tedy o stwierdzenie, czy – aby odwołać się do powyższych przykładów – nastąpiła redukcja poziomu lęku pac-

jentów depresyjnych pod wpływem stosowania wobec nich psychoterapii, czy też chodzi o sprawdzenie skuteczności danej techniki indoktrynacji zastosowanej wobec danej grupy osób przejawiających określoną postawę wobec jakiegoś obiektu. Na ową skuteczność wskazuje tu wielkość zmiany postawy zarejestrowana przez porównanie drugiego i pierwszego pomiaru jej natężenia w pierwszej grupie („indoktrynowanej”) jak i porównanie wielkości zmiany w drugiej grupie („relaksowej”). Badacz może też porównać dwie (i więcej) grupy traktowane w różnicowany sposób (np. dwa konkurencyjne sposoby leczenia, dwa konkurencyjne sposoby nauczania matematyki w pierwszej klasie szkoły podstawowej). Wówczas szuka on odpowiedzi na pytanie o to, który ze sposobów leczenia, czy sposobów nauczania jest efektywniejszy. Efektywniejszy jest ten, rzecz jasna, dla którego wielkość zmiany – w porównaniu z „konkurentem” – jest większa.

Badania zmian wymagają od badacza stosowanie przez niego dwukrotnego pomiaru zmiennej zależnej (Y). Pierwszy pomiar bywa określany mianem *pretestu* Y (w skrócie: Y_{pre}). Drugi pomiar bywa określany mianem *posttestu* Y (w skrócie Y_{post}). Stosuje się też określenia równoważne (por. Brzeziński 1984, s. 72): *pomiar początkowy* Y (w skrócie: Y_p) i *pomiar końcowy* Y (w skrócie: Y_k).

Stosowanie pretestu zmiennej zależnej w badaniach psychologicznych, socjologicznych, pedagogicznych, psychiatrycznych, a więc w badaniach, w których obiektem badanym jest człowiek, którego bada inny człowiek, rodzi swoiste problemy metodologiczne (a także etyczne – por. Brzeziński 1978), z którymi nie styka się, np. fizyk czy chemik. O tych właśnie problemach i próbach radzenia sobie z nimi będzie traktował niniejszy artykuł. Niejako przy okazji omówione będzie też zagadnienie postaci wyniku informującego badacza o zmianie interesującego go zjawiska.

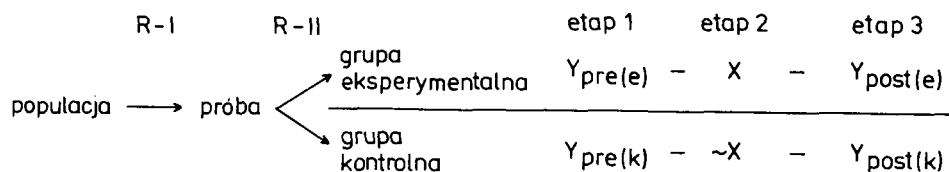
1. PODSTAWOWE PLANY EKSPERYMENTALNEGO BADANIA ZMIAN

PLAN DWUGRUPOWY Z PRZEPROWADZENIEM PRETESTU Y

Zgodnie z definicją planu eksperymentalnego (Brzeziński 1984, s. 62) wartości zmiennej niezależnej X (postępowanie eksperymentalne, zabieg eksperymentalny) przydzielone są losowo – zgodnie z zasadą randomizacji – do co najmniej dwóch grup porównawczych (eksperymentalnej i kontrolnej). I tak w grupie pierwszej, eksperymentalnej, badacz oddziałuje na postawy osób badanych, stosuje wobec pacjentów nowy, jeszcze nie sprawdzony lek, naucza dzieci zgodnie z regułami „nowej dydaktyki” (w skrócie: X). Z kolei osoby z grupy drugiej, kontrolnej, nie

poddane są zabiegom mającym doprowadzić do zmiany postaw, stosuje się wobec nich uznany, „klasyczny” lek, naucza się je zgodnie ze sprawdzonymi od dawna regułami „starej dydaktyki” (w skrócie: $\sim X$).

Badanie rozpoczyna się od losowego podziału osób (najlepiej też losowo dobranych z populacji) na dwie grupy porównawcze: eksperymentalną i kontrolną i przeprowadzenia pretestu Y . Następnie wobec osób z grupy eksperymentalnej i kontrolnej badacz stosuje zróżnicowane postępowanie (X vs $\sim X$). Badanie kończy się z chwilą wykonania posttestu Y . Schematycznie przedstawia to rys. 1.



Rys. 1. Podstawowy plan eksperymentalny, dwugrupowy, zakładający respektowanie zasad randomizacji: R-I i R-II

ZNACZENIE RANDOMIZACJI

Czy istotnie tak ważne jest losowe rozdzielanie osób badanych do grup porównawczych? Tak. Zwraçał już na to uwagę pionier nowoczesnego doświadczałnictwa i twórca analizy wariancji, Fisher (1925, 1935). On to wprowadził zasadę randomizacji jako podstawę przydzielania osób do grup. Na znaczenie randomizacji zwracają także uwagę: Oktaba (1966, s. 154-155), Lehmann (1968, s. 241), Hays (1973, s. 562-564) czy Linn i Slinde (1977, s. 132). Co najmniej dwa powody przemawiają za losowym rozdzielaniem osób badanych do grup porównawczych. Po pierwsze, randomizacja likwiduje działanie czynnika „selekcji” (wg Campbella i Stanleya 1967), który zakłóca trafność wewnętrzną planu eksperymentalnego. Mówiąc inaczej, jeżeli osoby nie będą wyznaczone losowo do grup porównawczych, to zawsze otwarte pozostanie pytanie: czy zaobserwowane różnice między przeciętnymi wartościami zmiennej zależnej (posttest Y) nie są spowodowane stronniczą selekcją osób do grupy eksperymentalnej i kontrolnej (np. dzieci bardziej uzdolnione matematycznie trafiły do klasy, w której nauczano wedle nowych zasad dydaktycznych, osoby bardziej podatne na psychoterapię trafiły do grupy leczonej psychoterapią, a nie do grupy leczonej tylko farmakologicznie itp.)? Aby zabezpieczyć się przed takimi zarzutami należy losować! Po drugie, stosowanie statystycznych testów istotności różnic oparte jest na

kilku założeniach. Jedno z nich obejmuje losową selekcję osób do grup porównawczych (por. Lehmann 1968, s. 241). Jeżeli wartość statystyki testowej wpada do obszaru odrzuceń hipotezy zerowej, to fakt ten możemy tłumaczyć bądź tym, co głosi hipoteza robocza (uzyskane istotne różnice między średnimi poziomami lęku w obu grupach można tłumaczyć w kategoriach zmiennej niezależnej X – „metoda leczenia pacjentów” – gdyż pacjentów z obu grup traktowano zróżnicowanymi metodami leczenia: „psychoterapia vs farmakoterapia”), bądź czynnikiem selekcji. Eliminując ten ostatni (via randomizacja!) możemy pozostać przy pierwszym sposobie tłumaczenia jako najbardziej prawdopodobnym. W przeciwnym wypadku oba sposoby są równo uprawnione jako potencjalne przyczyny niezachodzenia hipotezy zerowej. To, że w każdym warunkach (przy respektowaniu zasady randomizacji jak i przy jej pomijaniu) badacz raczej skłonny jest przyjmować „tłumaczenie” zarejestrowanej istotności różnic między średnimi wartościami zmiennej zależnej grupy eksperymentalnej i kontrolnej w kategoriach hipotezy roboczej niż w kategoriach selekcji osób do grup porównawczych świadczy o „przywiązaniu” się badacza do hipotezy roboczej (zjawisko to, pod nazwą: „efektu Rosenthala”, albo „efektu oczekiwań” opisane jest w: Rosenthal 1969, 1976; Rosenthal i Rubin 1978; Babad, Inbar i Rosenthal 1982).

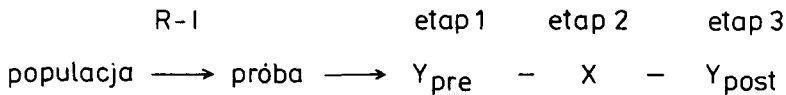
W jednej z prac (por. Brzeziński, Stachowski 1984, s. 114) rozróżniona została zasada randomizacji *I* od zasady randomizacji *II* (w skrócie: *R-I* i *R-II*). Zgodnie z pierwszą, badacz powinien losowo pobierać próbę z populacji, gdyż jedynie ten sposób gwarantuje jej reprezentatywność. Jest to również sposób na eliminację czynnika „interakcji selekcji i postępowania eksperymentalnego” zakłócającego trafność zewnętrzną planu eksperymentalnego (Campbell, Stanley 1967). Z kolei wyżej przedłożone racje będą racjami na rzecz randomizacji *II*.

Badanie przeprowadzone bez respektowania zasady randomizacji *I* dostarcza wyników, których nie można uznać za reprezentatywne dla określonej populacji. Nie można tedy uogólnić ich na daną populację. Wątpliwą też wartość mają rezultaty przeprowadzonego wnioskowania statystycznego (z użyciem testów istotności różnic). Takie badanie dostarcza wyników „prawdziwych” (i to przy założeniu niedziałania czynnika selekcji – respektowana zasada *R-II*) jedynie dla grupy, na której zostało ono przeprowadzone. Mówimy, że trafność ekologiczna takiego badania jest niska. Z kolei nierespektowanie zasady randomizacji *II* powinno uczulić badacza, w fazie interpretowania wyników, na to, że konkurencyjnym wyjaśnieniem zaobserwowanej istotności różnicy między grupami może być czynnik selekcji osób do tych właśnie grup. Plan eksperymentalny nie wymagający randomizacji *II* (takie plany eksperymentalne jednogrupowe, w których badacz wprowadza jakieś postępowanie eksperymentalne nie stosując równolegle grupy kontrolnej), to

plan – wedle przyjętej terminologii (por. Campbell, Stanley 1967; Brzeziński 1984, s. 62–63) – *quasi-eksperymentalny*. Taki plan ma zaniżoną, co najmniej, trafność wewnętrzną.

PLAN JEDNOGRUPOWY Z PRZEPROWADZENIEM PRETESTU Y

Innym, stosowanym przez badaczy planem jest, wspomniany już, plan jednogrupowy (por. rys. 2). Jest to plan quasi-eksperymentalny. Konkurencyjne do hipotezy roboczej wyjaśnienie zarejestrowanej istotności różnic:



Rys. 2. Plan quasi-eksperymentalny, jednogrupowy

„ $Y_{\text{post}} - Y_{\text{pre}}$ ” sprowadza się do zwrócenia uwagi na czynniki zakłócające trafność wewnętrzną takiego planu: czynnik „dojrzewania” (spontaniczne, wewnętrzne zmiany organizmu, psychiki związane z dojrzewaniem, naturalnymi procesami rozwoju fizjologicznego i psychicznego, samoistną rekonstrukcją organizmu pod wpływem samego „odstawienia” czynników szkodliwych, jak: alkohol, narkotyki, kontaktowanie się z grupami przestępczymi), czynnik historii (zmiany w otoczeniu osób badanych mogące wywrzeć wpływ na ich psychikę czy organizm, np. gwałtowne wydarzenia społeczne zachodzące w okresie dzielącym pretest Y od posttestu Y w badaniu funkcjonowania jakichś instytucji społecznych) – żeby ograniczyć się tylko do tych dwóch, najważniejszych czynników z listy zaproponowanej przez Campbella i Stanleya (1967) oraz Campbella (1969). Antidotum jest tylko jedno – dołączenie drugiej, kontrolnej grupy porównawczej. Zastrzeżenia powyższe nie dotyczą badania zmian spontanicznych, rozwojowych (wedle Krajewskiego 1977). Te jednak, nie mające charakteru eksperymentalnego tutaj nas nie interesują. Zresztą i w takich badaniach można dyskutować nad wpływem na zmienną zależną czynnika historii i nad próbą jego wyizolowania (por. strategię badań typu sekwencyjnego wypracowane w ramach tzw. „life-span psychology” – na ten temat: Baltes, 1978; Baltes, Brim, 1979; Baltes, Brim 1980; Baltes, Brim 1981; Baltes, Brim 1983).

PLANY WIELOGRUPOWE – DANE NIEZALEŻNE I ZALEŻNE

Rzecz jasna, efektywność badań przeprowadzonych wedle planów naszkicowanych na rysunkach 1 i 2 można zwiększyć przez wielokrotne stosowanie postępowania eksperymentalnego (o zróżnicowanym natężeniu) i posttestu Y w czasie, np. podawanie dawki leku co jakiś czas i dokonywanie pomiaru poziomu lęku, wielokrotny pomiar natężenia postawy w trakcie oddziaływania socjotechnicznego. Takie badania dostarczają nam informacji o dynamice interesującego nas zjawiska i umożliwiają – via specjalistyczne opracowanie statystyczne – zanalizowanie trendu rejestrowanych zmian (liniowy *vs* krzywoliniowy). I tak możliwe jest przeprowadzenie parametrycznej analizy trendu (por. Winer 1971) oraz nieparametrycznej – z wykorzystaniem testu Friedmana (por. Brzeziński, Maruszewski 1981). Efektywność badania można też zwiększyć rezygnując z planów: „0-1” (jedna grupa eksperymentalna – jedna grupa kontrolna) na rzecz planów wielogrupowych (dane niezależne). W tych ostatnich zmienna X nie jest na „siłę” sprowadzana do postaci dychotomicznej, ale może przybierać więcej wartości (sprowadzanie zmiennej wielowartościowej do postaci dwuwartościowej związane jest z utratą, bywa że dość znaczną, informacji o powiązaniu owej zmiennej ze zmienną zależną – por. Cohen 1983). Umożliwia to także analizę trendu między zmiennymi X i Y . Podstawy metody analizy trendu opartej na modelu analizy wariancji (ANOVA) przedstawiam w: Brzeziński (1985), a odmianę nieparametryczną – wychodzącą z testu Kruskala – Wallisa omawia praca: Brzeziński i Maruszewski (1981).

2. KLASYCZNA ANALIZA STATYSTYCZNA WYNIKÓW EKSPERYMENTALNEGO BADANIA ZMIAN

PORÓWNANIA DWÓCH GRUP NIEZALEŻNYCH

Pierwszy moment, w którym badacz styka się z problemem pretestu zmiennej Y , to moment, w którym sprawdza on równoważność – pod względem „wyjściowego” natężenia zmiennej zależnej – obu grup: eksperymentalnej i kontrolnej. Chodzi o sprawdzenie, czy zachodzi równość: „ $\bar{Y}_{pre(e)} \stackrel{s}{=} \bar{Y}_{pre(k)}$ ”. Litera „s” pod znakiem równości wskazuje na to, że nie oczekujemy dosłownej równości wyników grupy eksperymentalnej i kontrolnej, ale że oczekujemy, iż zarejestrowana różnica między, np. średnimi wartościami zmiennej zależnej w obu grupach będzie na tyle mała, że zastosowany przez badacza test istotności różnic, np. test t , nie upoważni go do odrzucenia hipotezy zerowej postaci: $H_0: \mu_e = \mu_k$ na rzecz hipotezy roboczej postaci: $H_1: \mu_e \neq \mu_k$.

Jeżeli przyjmiemy za obowiązujący podział testów istotności różnic

wedle kryterium skali pomiarowej zmiennej zależnej (za: Stevens 1951), to badacz może zastosować jakiś z najbardziej znanych testów parametrycznych (na temat ANOVA por. Brzeziński, Stachowski 1984; Winer 1971; z kolei ANCOVA omówiona jest przez Winer 1971; MANOVA i MANCOVA – por. Timm 1975; Cooley i Lohnes 1971; testy: *z* i *t* – np. Blalock 1975) i nieparametrycznych (duży wybór znajdzie Czytelnik w: Hollander i Wolfe 1973; Domański 1979; Siegel 1956). Będą to tylko testy z grupy testów przeznaczonych dla porównań dwóch (i większej liczby) grupy niezależnych (por. tabela 1).

Tabela 1

Najbardziej znane testy istotności różnic dla porównań grup niezależnych – do oceny zachodzenia warunku: „ $\bar{Y}_{pre(e)} = \bar{Y}_{pre(k)}$ ” w planach eksperymentalnych

Skala pomiaru zmiennej Y	Test istotności różnic
interwałowa i ilorazowa	test F – ANOVA, ANCOVA, MANOVA, MANCOVA (dla <i>k</i> grup) test <i>z</i> test <i>t</i>
porządkowa	test Kołmogorowa-Smirnowa test Manna-Whitneya test Walda-Wolfowitza test Kruskala-Wallis (dla <i>k</i> grup)
nominalna	test chi-kwadrat dla tabel 2 x 2 test chi-kwadrat dla tabel <i>r</i> x <i>c</i> (dla <i>k</i> grup) test dokładnego prawdopodobieństwa – Fishera dla tabel 2 x 2

ROLA PRETESTU Y

Stosowanie pretestu Y w planach eksperymentalnych, dwugrupowych (i wielogrupowych) spełnia ważne funkcje. Po pierwsze, badacz uzyskuje informacje o poziomie zmiennej zależnej przed wprowadzeniem postępowania eksperymentalnego. Przez przyrównanie tego poziomu do poziomu zmiennej zależnej, określonego po ustaniu działania postępowania eksperymentalnego, badacz może określić wielkość zmiany jaka zaszła u konkretnej osoby badanej pod wpływem oddziaływania na Y zmiennej X. Po drugie, określenie poziomu wyjściowego zmiennej Y w obu grupach jednocześnie pozwoli badaczowi rozwiązać ewentualne wątpliwości co do „równego startu” grup – eksperymentalnej i kontrolnej. Po trzecie, badacz może określić relatywną wielkość przeciętnej zmiany zarejestrowanej w grupie eksperymentalnej wobec wielkości przeciętnej zmiany zaobserwowanej w grupie kontrolnej. Czym innym bowiem jest szukanie odpowiedzi na pytanie o wpływ zastosowanego środka psychostymulacyjnego na zwiększenie liczby poprawnych odpowiedzi w teście

badającym koncentrację uwagi, a czym innym jest wybór jednej z dwóch metod terapii jakiegoś zaburzenia psychicznego (farmakoterapia vs farmakoterapia plus psychoterapia, znany lek vs nowy lek), czy jednej z dwóch metod nauczania (klasyczna vs nowa). W pierwszym przypadku bowiem stwierdzenie przyrostu liczby poprawnych rozwiązań w grupie eksperymentalnej, po zakończeniu eksperymentu, pozwoli badaczowi na przyjęcie hipotezy roboczej ($H_1: \mu_{\text{post}(e)} - \mu_{\text{pre}(e)} > 0$) i nieodrzućanie hipotezy zerowej w odniesieniu do grupy kontrolnej ($H_0: \mu_{\text{post}(k)} - \mu_{\text{pre}(k)} = 0$). Z kolei w drugim przypadku badacz musi założyć, że i w grupie kontrolnej należy się spodziewać wystąpienia zmiany, tyle tylko, że mniejszej niż w grupie eksperymentalnej. Oczekiwanie zatem, że pacjenci leczenia tylko farmakologicznie, czy uczniowie nauczani tradycyjnie, ani nie poprawią stanu swojego zdrowia, ani nie zwiększy się zasób ich wiedzy (zgodnie z hipotezą „ $\mu_{\text{post}(k)} - \mu_{\text{pre}(k)} = 0$ ” – ta hipoteza miała sens w pierwszym przypadku) jest niczym nie uzasadnione. I farmakoterapia jest w jakimś stopniu skuteczna i tradycyjne nauczanie także. Zatem badacz musi – w takich i podobnych przypadkach – inaczej podejść do formułowania hipotez i do ich oceny. Hipoteza robocza powinna przyjąć postać, w której „wpisany” będzie relatywizm porównań. Taką prawidłową postać ma hipoteza: $H_1: \bar{d}_e > \bar{d}_k$, gdzie: $\bar{d}_e = \mu_{\text{post}(e)} - \mu_{\text{pre}(e)}$, $\bar{d}_k = \mu_{\text{post}(k)} - \mu_{\text{pre}(k)}$ przy jednoczesnym założeniu słuszności $H_0: \mu_{\text{pre}(e)} = \mu_{\text{pre}(k)}$. Mówiąc inaczej, zarówno w grupie eksperymentalnej jak i w kontrolnej oczekujemy wystąpienia zmiany, tyle tylko, że będzie ona relatywnie większa w grupie eksperymentalnej niż w kontrolnej.

PORÓWNANIE DWÓCH GRUP ZALEŻNYCH

Oceny istotności zmiany typu: „ $\bar{Y}_{\text{post}} \underset{S}{\geq} \bar{Y}_{\text{pre}}$ ” dokonuje się za pomocą testów istotności różnic przeznaczonych dla porównań grup zależnych.

Tabela 2

Najbardziej znane testy istotności różnic dla porównań grup zależnych – do oceny zachodzenia warunku: „ $\bar{Y}_{\text{post}(e)} \underset{S}{=} \bar{Y}_{\text{pre}(e)}$ ”

Skala pomiaru zmiennej Y	Test istotności różnic
interwałowa i ilorazowa	test F – ANOVA, ANCOVA, MANOVA, MANCOVA (dla k grup) test z test t
porządkowa	test Wilcoxona test Friedmana (dla k grup)
nominalna	test McNemara dla tabel 2 x 2 test dwumianowy test Cochrańa (dla k grup)

Najważniejsze z nich – parametryczne i nieparametryczne – przytacza tabela 2. Ich opis znajdzie Czytelnik w tych samych pracach, które przytoczyłem przy wymienianiu testów dla grup niezależnych.

3. PRETEST ZMIENNEJ ZALEŻNEJ JAKO ŹRÓDŁO ARTEFAKTÓW

PIĘĆ RODZAJÓW PORÓWNAŃ

W uwagach przedstawionych w poprzednim punkcie wystąpiły trzy rodzaje porównań, które są dla badacza podstawą do podjęcia decyzji odnośnie do testowanych hipotez:

a) porównanie pretestów Y obu grup, eksperymentalnej i kontrolnej, np. $Y_{pre(e)} \stackrel{S}{=} Y_{pre(k)}$;

b) porównanie posttestu Y z pretestem Y w danej grupie, eksperymentalnej lub kontrolnej, np. $Y_{post(e)} \stackrel{S}{\geq} Y_{pre(e)}$;

c) porównanie różnic (d) utworzonych z porównania posttestu Y z pretestem Y w obu grupach, eksperymentalnej i kontrolnej – $d_e \stackrel{S}{\geq} d_k$. Stosowane są jeszcze dwa rodzaje porównań:

d) porównanie wyników resztowych (Y_{res}) utworzonych z porównania zaobserwowanych wyników posttestu Y z wynikami przewidywanymi posttestu Y obliczonymi z równania regresji liniowej, w którym rolę predyktora odgrywają wyniki pretestu Y ; w obu grupach, eksperymentalnej i kontrolnej, oczekuje się, że: $Y_{res(e)} \stackrel{S}{\geq} Y_{res(k)}$;

e) porównanie posttestów Y (w planach dwu- i wielogrupowych) postaci: $Y_{post(e)} \stackrel{S}{\geq} Y_{post(k)}$; ten rodzaj porównań, z uwagi na niewystępowanie w nim wyników pretestu Y nie będzie nas dalej interesował.

Zajmiemy się teraz trzecim i czwartym rodzajem porównań (c,d) z uwagi na osobliwość wyników – zapisanych w postaci różnicy – które się na nie składają.

WYNIKI RÓŻNICOWE

Różnice indywidualne obliczane dla każdej osoby z grupy eksperymentalnej i kontrolnej są, jak to już powiedzieliśmy, podstawą do testowania hipotez zerowych postaci: $H_0: \bar{d}_e = \bar{d}_k$. Wyniki różnicowe dla osób z grupy eksperymentalnej i kontrolnej przyjmują postać:

<i>grupa eksperymentalna</i>	<i>grupa kontrolna</i>	
$d_{1e} = Y_{post(1e)} - Y_{pre(1e)}$	$d_{1k} = Y_{post(1k)} - Y_{pre(1k)}$	
$d_{2e} = Y_{post(2e)} - Y_{pre(2e)}$	$d_{2k} = Y_{post(2k)} - Y_{pre(2k)}$	
.....	
$d_{ne} = Y_{post(ne)} - Y_{pre(ne)}$	$d_{nk} = Y_{post(nk)} - Y_{pre(nk)}$	
<hr/>	<hr/>	
$d_{.e}$	$d_{.k}$	<i>suma</i>
$\bar{d}_{.e}$	$\bar{d}_{.k}$	<i>średnia</i>

Podstawowym przeciwskazaniem do posługiwania się wynikami różnicowymi jest fakt zachodzenia ujemnej korelacji między wynikami pretestu Y i wynikami różnicowymi. Na tę „wadę” wskazywali już Bereiter (1963), Thorndike (1966) czy Linn i Slinde (1977). Przypomnijmy, iż korelacja między wynikami pretestu Y i wynikami różnicowymi [$r_{(pre)(d)}$] wyrażona jest wzorem (Cohen, Cohen 1975, s. 380):

$$r_{(pre)(d)} = \frac{r_{(pre)(post)}S_{(post)} - S_{pre}}{\sqrt{S_{pre}^2 + S_{post}^2 - 2r_{(pre)(post)}S_{pre}S_{post}}} \quad (1)$$

s – odchylenie standardowe

O wielkości owej korelacji – jak nietrudno zauważyć analizując licznik wzoru (1) – decyduje wielkość korelacji między wynikami pretestu Y i posttestu Y . Konsekwencją ujemnej korelacji między wynikami pretestu Y i posttestu Y jest to, że relatywnie osoby o niskich wynikach pretestu Y będą miały raczej wyższe wyniki różnicowe, a osoby o wysokich wynikach pretestu Y będą miały raczej niskie wyniki różnicowe. Zjawisko to ma także i dalsze konsekwencje w postępowaniu selekcyjnym. Zatem, jeżeli badacz zdecyduje się dobierać do grup badawczych osoby o wysokich wynikach różnicowych (uzyskanych na podstawie wstępnych badań selekcyjnych), to doprowadzi do nadreprezentacji w badanej próbie osób o niskich wynikach pretestu Y . Posługiwanie się tedy wynikami różnicowymi może stanowić dodatkowe źródło artefaktów w badaniach społecznych.

Lord (1963) wskazał na jeszcze jedną „wadę” wyników różnicowych, na ich relatywnie niską rzetelność. Jeżeli przez rzetelność będziemy rozumieli, tak jak Gulliksen (1950) oraz Lord i Novick (1968), m.in. stabilność bezwzględną [termin zaproponowany w polskiej literaturze psychometrycznej przez Choynowskiego (1971, s. 85)], której miarą jest współczynnik korelacji: „test – retest” (korelacja wyników dwukrotnego pomiaru danej cechy przeprowadzonego za pomocą tego samego testu w jakimś odstępie czasu – r_t), to rzetelność wyników różnicowych (r_{dd}) wyrażona jest wzorem:

$$r_{dd'} = \frac{r_{(pre)(pre')}S_{pre}^2 + r_{(post)(post')}S_{post}^2 - 2r_{(pre)(post)}S_{pre}S_{post}}{S_{pre}^2 + S_{post}^2 - 2r_{(pre)(post)}S_{pre}S_{post}} \quad (2)$$

$r_{(pre)(pre')}$ – współczynnik rzetelności pretestu Y ,

$r_{(post)(post')}$ – współczynnik rzetelności posttestu Y ,

$r_{(pre)(post)}$ – współczynnik korelacji pretestu Y i posttestu Y ,

s – odchylenie standardowe.

Jeżeli przypatrzymy się uważnie wzorowi (2), to zauważymy, iż $r_{dd'}$ zależy od wielkości współczynnika korelacji $r_{(pre)(post)}$. Im wyższy jest współczynnik korelacji między wynikami pretestu Y i posttestu Y , tym

niższa jest rzetelność wyników różnicowych. Dla przykładu – jeżeli rzetelność pretestu Y i posttestu Y jest taka sama i wynosi: 0,90, to przy korelacji „pretest-posttest” równej 0,50 rzetelność wyników różnicowych wynosi: 0,80, ale przy korelacji „pretest-posttest” równej 0,00 ta rzetelność jest już alarmująco niska i wynosi tylko: 0,50 (dane zaczerpnięte z Tab. 1, s. 123, Linn, Slinde 1977). Zatem jest to drugi powód, aby nie posługiwać się wynikami różnicowymi w ważnych (brzemiennych w skutki dla osób badanych lub instytucji społecznych) badaniach psychologicznych, socjologicznych itp. Także z powodu niskiej rzetelności wyniki różnicowe mogą stać się źródłem artefaktów w badaniach prowadzonych na terenie nauk społecznych. Osobliwość związków między korelacją pretestu Y i posttestu Y , samymi wynikami różnicowymi i rzetelnością wyników różnicowych prowadzi do paradoksu pomiarowego, jeśli chodzi o pomiar zmian. Paradoks ten opisany jest przez Overalla i Woodwarda (1975). Jak wiadomo, istotność różnic między średnimi wynikami posttestu Y i pretestu Y można ocenić za pomocą testu t (dla danych zależnych): $t = (\bar{Y}_{\text{post}} - \bar{Y}_{\text{pre}}) : s_d$, gdzie: s_d – nieobciążony estymator odchylenia standardowego z próby, $s_d^2 = s_d^2 : n$. Z kolei: $s_d^2 = s_{\text{pre}}^2 + s_{\text{post}}^2 - 2r_{(\text{pre})(\text{post})}$. Widać tedy wyraźnie, że wartość statystyki testowej „ t ” będzie wzrastała wraz ze wzrostem współczynnika korelacji $r_{(\text{pre})(\text{post})}$, gdyż tym samym będzie malała wartość s_d (mianownik testu t). Jednocześnie wzór (2) „pokazuje”, że wraz ze wzrostem korelacji: „pretest – posttest” maleje rzetelność wyników różnicowych. I to jest istotą owego paradoksu opisanego przez Overalla i Woodwarda.

WYNIKI RESZTOWE

Badaczom zorientowanym „statystycznie” na model analizy regresji liniowej bliższa będzie konstrukcja wyniku informującego o zmianie oparta na tzw. resztach (ang. *residuals* – por. Draper, Smith 1973, s. 108 i nast. Powstaje on przez odjęcie wyniku zaobserwowanego posttestu Y od wyniku przewidywanego posttestu Y (oznaczanego jako: Y'_{post}):

$$\begin{aligned} Y_{\text{res}} &= Y_{\text{post}} - Y'_{\text{post}} \\ Y' &= bY_{\text{pre}} + a \end{aligned}$$

Wyniki resztowe są bardziej rzetelne niż wyniki różnicowe. Jednakże i one (ich rzetelność) zależą od stopnia korelacji pretestu Y z posttestem Y . Mimo wszystko, ta zależność w przypadku wyników resztowych jest mniejsza.

Współczynnik rzetelności wyników resztowych dany jest wzorem (O'Connor 1972):

$$r_{(\text{res})(\text{post}')} = \frac{r_{(\text{post})(\text{post}')} - r_{(\text{pre})(\text{post})}^2 [2 - r_{(\text{pre})(\text{pre}')}]}{1 - r_{(\text{pre})(\text{post})}^2} \quad (3)$$

Wyniki resztowe są przydatne zwłaszcza wówczas, gdy korelacja pretestu Y z posttestem Y jest relatywnie większa od współczynników rzetelności pretestu Y i posttestu Y.

Dotychczasowe nasze rozważania nad różnymi postaciami wyniku informującego badacza o wystąpieniu zmiany skupione były wokół zagadnień natury formalnej i statystycznej. Można by na tym poprzestać, gdyby były to wyniki spoza nauk społecznych. Trzeba jednak pamiętać, że psycholog, socjolog, psychiatra, pedagog dokonując pomiaru zmiennej zależnej (pretest!), dokonują go wchodząc w interakcje z osobą badaną. Osoba badana nadaje aktowi pomiaru określony sens. Badany, jak to pokazał Orne (1962, 1969, 1970), stara się na podstawie analizy różnych wskaźników, których dostarcza mu postępowanie badawcze (także stosowane testy inteligencji, skale postaw, kwestionariusze osobowości itp.), „rozszyfrować” hipotezy robocze i w zależności od tego, czy udział w badaniu jest dla niego korzystny (np. osoba badana otrzymuje za udział w badaniu wynagrodzenie) czy szkodliwy (badanie może, na przykład, ujawnić „słabe strony” osoby badanej) będzie z badaczem współpracował (działając na rzecz wyobrażonej hipotezy) lub odmówi współpracy. Badanie inteligencji, osobowości, opinii wywołuje u badanych lęk przed oceną (Rosenberg 1965, 1969) zwłaszcza wówczas, gdy pomiar dotyczy osiowych elementów obrazu siebie samego, a osoby badane cechuje niestabilna i zaniziona samoocena. Także badanie postaw uwrażliwia osoby badane i sprawia, że zachowują się one w trakcie badania nienaturalnie.

UWRAZLIWIJĄCY CHARAKTER PRETESTU Y

Stosując pretest zmiennej zależnej ingerujemy w delikatny układ badawczy zmieniając go (w jakim stopniu?). Osoby z badanej próby – po wykonaniu pretestu Y – różnią się już od osób z populacji, która nie jest poddawana pretestowi. Jeżeli sprawdzamy skuteczność jakiejś techniki zmiany postaw, to po przeprowadzeniu pretestu Y część osób badanych zacznie werbalizować swój stosunek do obiektów, o których była mowa w skali postaw służącej do przeprowadzenia pretestu Y. Może tedy być pretest Y dodatkowym źródłem wariacji zmiennej zależnej (posttestu Y). Może być i tak, że postępowanie eksperymentalne wejdzie w interakcję z pretestem Y. Wreszcie nie można wykluczyć i takiego wariantu wyjaśnienia zaobserwowanej zmiany postaw, który dopuszcza pretest Y jako jedyne źródło wariacji wyjaśnionej zmiennej zależnej. Na takie ewentualności wskazują Campbell i Stanley (1967) mówiąc o czynniku „testowania” jako o czynniku zakłócającym trafność wewnętrzną planu eksperymentalnego oraz zwracają uwagę na ograniczenia w generalizowaniu wniosków z próby na populację (jeżeli próba była przebadana za

pomocą narzędzia uwrażliwiającego osoby badane) mówiąc o czynniku zakłócającym trafność zewnętrzną planu eksperymentalnego, tj. o „reaktywnym albo interakcyjnym efekcie testowania”. Czy można tedy bez zastrzeżeń uogólnić wyniki takiego badania na populację, która nie podlega pretestowi Y? Rodzi się tedy następujący problem: w jaki sposób poddać efektywnej kontroli uwrażliwiający charakter pretestu Y? Mówiąc językiem bardziej technicznym: jak określić (i oddzielić) wielkość wariacji składowej całkowitej wariacji zmiennej zależnej (posttestu Y), której źródłem (prawdopodobnie) jest pretest Y i jego interakcja z postępowaniem eksperymentalnym? O tym jednak będziemy mówili w następnym punkcie niniejszej pracy. Rodzi się w tym miejscu analogia do słynnej w fizyce zasady nieoznaczoności – Heisenberga. W mikroświecie, który bada fizyk kwantowy też zdarza się naruszenie „delikatnego” układu cząsteczkowego przez stosowaną przez fizyka metodę ustalenia położenia cząsteczki, czy ustalenia jej pędu. Rzeczywistość „zmierzona” staje się inną rzeczywistością (por. Ossowski 1967, s. 255-257). W naukach społecznych dokonując pomiaru też ingerujemy w delikatną tkankę psychologiczną czy społeczną zmieniając ją i w konsekwencji kreując „nowy” byt, którego relacje do „starego” bytu są badaczowi nieznane. Zostaje tedy Hamletowskie pytanie: mierzyć albo nie mierzyć ...?

PROBLEM RÓWNOLEGŁOŚCI PRETESTU Y I POSTTESTU Y

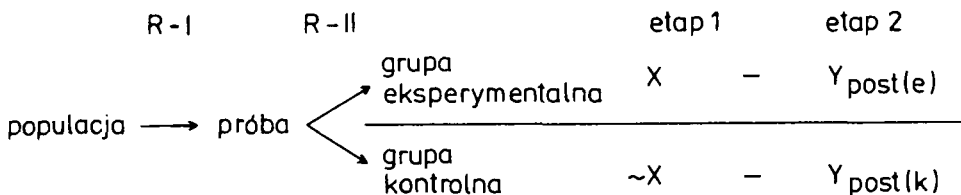
Kończąc nasze rozważania o uwrażliwiający osoby badane charakterze pretestu Y zwróćmy jeszcze uwagę na jedno zagadnienie. Aby trafnie dokonać oceny wielkości zmiany w okresie dzielącym pretest Y od posttestu Y musimy posłużyć się w obu przypadkach tym samym narzędziem, tj. tym samym testem inteligencji, tą samą skalą postaw, tym samym testem wiadomości itp. Jednakże trudne pytanie z testu wiadomości w pierwszym badaniu (pretest Y) staje się łatwe w drugim badaniu (posttest Y) niezależnie od skuteczności postępowania eksperymentalnego (osoba badana może je zapamiętać i poszukać prawdziwej odpowiedzi w podręczniku, czy spytać się osoby kompetentnej). Próbą wyjścia, ale pod względem psychometrycznym trudną, a czasem praktycznie niewykonalną, jest posługiwanie się dwiema formami równoległymi tego samego narzędzia (por. Gulliksen 1950, s. 173-192). Wówczas za pomocą jednej formy przeprowadzamy pretest Y, a za pomocą drugiej – posttest Y. Problem konstrukcji form równoległych testu psychologicznego, w celu przeprowadzenia pretestu Y i posttestu Y wydaje się być jeszcze rozwiązywalny (w sensie praktycznym) w przypadku planów eksperymentalnych, dwugrupowych. Natomiast sytuacja staje się beznadziejna, gdy badacz zamierza przeprowadzić badanie wg schematu wielogrupowego. Aby wyjść obronną ręką z tego dylematu, należy szukać (konstruować)

takich metod pomiaru, które wywołują jak najmniejsze uwrażliwienie osób badanych. Przykładowo, psycholog może o poziomie lęku osób badanych wnosić na podstawie wyników kwestionariusza *STAI*-Spielberga, albo na podstawie wskaźników obiektywnych (leżących poza zasięgiem świadomości osób badanych) takich jak pomiary elektrodermalne w rodzaju *GSR*, *EKG*, *EEG* itp. (por. Geras 1985). Jeżeli mamy do wyboru kwestionariusz samooceny (wypełnia go osoba badana) lub kwestionariusz oceny (wypełnia go obserwator, np. obserwacyjna skala do oceny stopnia nasilenia objawów depresyjnych w układzie Hamiltona) i obojętne jest z punktu widzenia trafności procedury operacjonalizacji, który z tych kwestionariuszy wybierzemy, to powinniśmy zdecydować się raczej na kwestionariusz oceny, gdyż nie będziemy mieli do czynienia z problemem uwrażliwienia osób badanych z powodu zastosowanej formy pretestu Y.

4. SPOSOBY RADZENIA SOBIE Z PRETESTEM ZMIENNEJ ZALEŻNEJ W PLANACH EKSPERYMENTALNYCH I QUASI-EKSPERYMENTALNYCH

PLANY Z CAŁKOWITĄ LUB CZĘŚCIOWĄ ELIMINACJĄ PRETESTU Y.

Jeżeli badacz nie zechce skorzystać z wyżej przedstawionej rady i nie zastąpi (bo nie może) narzędzia uwrażliwiającego osoby badane takim, które owej negatywnej właściwości nie posiada, to może przeprowadzić badanie wg planu, w którym całkowicie wyeliminowany jest pretest Y. Plan ten przedstawiony jest na rys. 3. Jest to, jak nietrudno zauważyć wariant planu z rys. 1 powstały przez „obcięcie” etapu pretestu Y. Plany

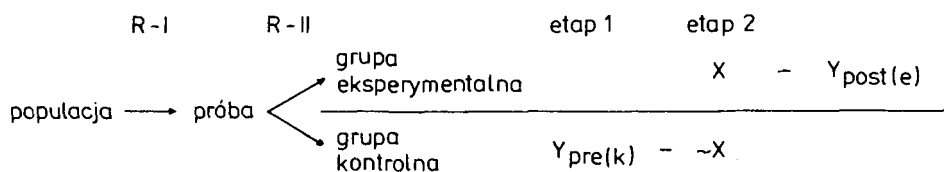


Rys. 3. Plan eksperymentalny z wyeliminowaniem pretestu Y

takie omawiane są w literaturze przedmiotu (Campbell, Stanley 1967; Muszyński 1971, s. 266-267; Brzeziński 1984, s. 73). Uważam, że są one całkowicie nieprzydatne do rozwiązywania problemów badania zmian.

Chodzi nam wszak o to – przypomnijmy – aby możliwe było dokonanie oceny tego, jak wielka zaszła zmiana w zachowaniu się osób badanych z grupy eksperymentalnej (w stosunku do osób z grupy kontrolnej) pod wpływem postępowania eksperymentalnego. Nie można tedy, zastąpić porównania: „ $d_e \geq d_k$ ”, czy „ $Y_{\text{post}(e)} \geq Y_{\text{pre}(e)}$ ” porównaniem: „ $Y_{\text{post}(e)} \geq Y_{\text{post}(k)}$ ”.

Inną próbą ominięcia trudności związanych z przeprowadzaniem pretestu Y , ale nie rezygnacji z niego, jest przeprowadzenie badania wedle planu przedstawionego na rys. 4. Jak widzimy pretest Y przeprowa-

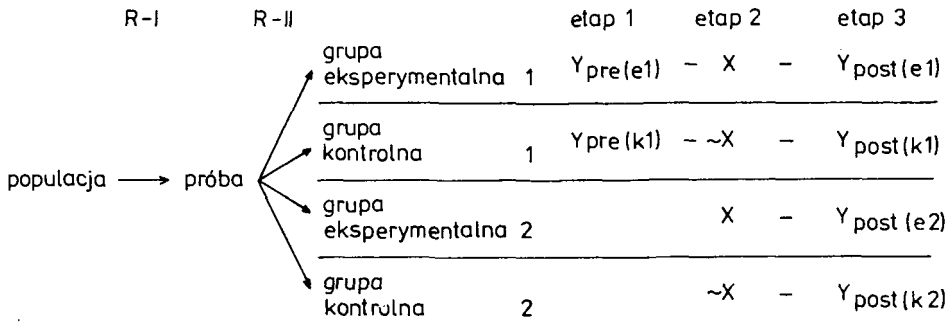


Rys. 4. Plan quasi-eksperymentalny – pretest Y i posttest Y przeprowadzone na różnych grupach

dzamy tylko w grupie kontrolnej, w stosunku do której nie stosujemy dalej ani postępowania eksperymentalnego ani posttestu Y . Z kolei w grupie eksperymentalnej wprowadzamy postępowanie eksperymentalne i wykonujemy posttest Y . Testujemy hipotezę roboczą sprawdzając czy zachodzi porównanie: „ $Y_{\text{post}(e)} \geq Y_{\text{pre}(k)}$ ”. Plan taki opisany jest m.in. przez: Campbella i Stanleya (1967), Muszyńskiego (1971, s. 268), czy Brzezińskiego (1984, s. 87). Plan ten rodzi jednak inną trudność. Jak sprawdzić, że grupa kontrolna w momencie przeprowadzania pretestu Y była „taka sama”, jak grupa eksperymentalna?

PLAN SOLOMONA – PEŁNA KONTROLA PRETESTU Y

Nie rezygnując z przeprowadzenia pretestu Y można go jednak poddać efektywnej kontroli. Będzie ona polegała na rozbiciu wariacji posttestu Y na wariacje składowe: wprowadzoną przez postępowanie eksperymentalne (X), wprowadzoną przez pretest Y (Y_{pre}) oraz wprowadzoną przez interakcję postępowania eksperymentalnego z pretestem Y . Będzie to możliwe, jeżeli badanie będzie przeprowadzone wedle planu zaproponowanego przez Solomona (1949). Jest to udana próba poddania kontroli pretestu Y , bez rezygnowania z niego. Plan ten, przedstawia go rys. 5, nadal jest rekomendowany w literaturze przedmiotu (por. np. Campbell i Stanley 1967); Czabała i in., 1973, Brzeziński 1975, 1984, s. 74).



Rys. 5. Czterogrupowy plan Solomona – dwie grupy eksperymentalne i dwie grupy kontrolne

Plan Solomona powstał przez połączenie dwóch planów – przedstawionych na rysunkach 1 i 2. Obejmuje on dwie grupy eksperymentalne i dwie grupy kontrolne. Umożliwia on pełną kontrolę wszystkich czynników zakłócających trafność wewnętrzną i zewnętrzną planu eksperymentalnego. Przeprowadzając serię porównań inter- i intragrupowych (za pomocą jakiegoś testu istotności różnic – por. tabelę 1 i 2) można odpowiedzieć na pytanie o skuteczność postępowania eksperymentalnego, wyrażoną wielkością zaobserwowanej zmiany, a także na pytanie o uwrażliwiający osoby badane wpływ pretestu Y (porównania te przedstawia Brzeziński 1984, s. 75).

Gdyby się jednak okazało, że pretest Y działał uwrażliwiająco na osoby badane, to – pozostając przy tych dwugrupowych analizach statystycznych – nie możemy określić „czystego” wpływu postępowania eksperymentalnego na posttest Y od pozostałych wpływów związanych z przeprowadzonym pretestem Y. Aby mimo wszystko dokonać oceny stopnia zagrożenia wyników posttestu Y przez przeprowadzenie pretestu Y musimy się odwołać do bardziej zaawansowanych modeli statystycznych. Zaczniemy wpierw od modelu analizy wariancji (ANOVA), za pomocą którego – jak zobaczymy – możliwe jest statystyczne opracowanie wyników badania eksperymentalnego przeprowadzonego zgodnie z planem Solomona.

MODEL ANOVA A KONTROLA PRETESTU Y W PLANIE SOLOMONA¹

W badaniu eksperymentalnym przeprowadzonym wg planu Solomona całkowitą wariancję zmiennej zależnej (posttestu Y) można rozbić na następujące wariancje składowe:

¹ W tym punkcie niniejszej pracy częściowo oparłem się na wcześniej napisanym artykule: Brzeziński (w druku – a)

- a – $\text{var}(Y_{\text{post}} | X)$ – wariancja składowa wyjaśniona wpływem na Y postępowania eksperymentalnego;
- b – $\text{var}(Y_{\text{post}} | Y_{\text{pre}})$ – wariancja składowa wyjaśniona wpływem na Y pretestu Y ;
- c – $\text{var}(Y_{\text{post}} | XY_{\text{pre}})$ – wariancja składowa wyjaśniona wpływem na Y interakcji postępowania eksperymentalnego z pretestem Y ;
- d – $\text{var}(Y_{\text{post}} | \text{reszta})$ – wariancja składowa, tzw. „wariancja resztowa”, której źródłem są zmienne niezależne nie kontrolowane przez badacza.

Wpisując wyniki posttestów Y (pochodzące od czterech grup porównawczych występujących w planie Solomona) do tabeli takiej jak na rys. 6 przeprowadzamy analizę wariancji w układzie dwuczynnikowym mo-

		Pretest Y	
		Y_{pre}	$\sim Y_{\text{pre}}$
Postępowanie eksperymentalne	X	$Y_{\text{post}(e1)}$	$Y_{\text{post}(e2)}$
	$\sim X$	$Y_{\text{post}(k1)}$	$Y_{\text{post}(k2)}$

Rys. 6. Tabela wyników w układzie dwuczynnikowej analizy wariancji – do danych uzyskanych wg planu Solomona (rys. 5)

del efektów stałych (por. Brzeziński, Stachowski 1984, s. 190 i nast). Dane z sumarycznej tabeli analizy wariancji posłużą nam do przeprowadzenia analizy wariancji składowych (a – d). Metoda ta, zaproponowana przez Crumpa (1951) została wykorzystana przez wielu autorów do opracowania ogólniejszych reguł i konkretnych wzorów na procentowo ujęte wskaźniki udziału poszczególnych wariancji składowych w całej wariancji posttestu Y . Nie wchodząc w istotę tej metody i jej uzasadnienie (w tej sprawie patrz: Endler 1966; Halderson, Glasnapp 1971; Sechrest, Yeaton 1982; Brzeziński 1983) podam od razu „gotowe” wzory obliczeniowe.

Wariancję składową (a) – (d) obliczamy wg wzoru:

$$\hat{\text{var}}_X = \frac{(p-1)(MS_X - MS_{\text{reszta}})}{npq}$$

$$\hat{\text{var}}_{Y_{\text{pre}}} = \frac{(q-1)(MS_{Y_{\text{pre}}} - MS_{\text{reszta}})}{npq}$$

$$\hat{\text{var}}_{XY_{\text{pre}}} = \frac{(p-1)(q-1)(MS_{XY_{\text{pre}}} - MS_{\text{reszta}})}{npq}$$

$$\hat{\text{var}}_{\text{reszta}} = MS_{\text{reszta}}$$

oznaczenia: p – liczba poziomów zmiennej X (tu: $p = 2$),
 q – liczba poziomów zmiennej Y_{pre} (tu: $q = 2$),
 n – liczba osób badanych w jednej grupie (tu: $n = n_1 = n_2 = n_3 = n_4$),
 MS – średnie kwadraty (wzięte z sumarycznej tabeli ANOVA).

Oszacowanie całkowitej wariancji Y_{post} dane jest wzorem:

$$\hat{\text{var}}_{\text{całkowita}} = \hat{\text{var}}_X + \hat{\text{var}}_{Y_{\text{pre}}} + \hat{\text{var}}_{XY_{\text{pre}}} + \hat{\text{var}}_{\text{reszta}}$$

Chcąc dokonać procentowego oszacowania udziału danej wariancji składowej – np. wyjaśnianej efektem pretestu Y – w wariancji całkowitej Y_{post} musimy odnieść $\hat{\text{var}}_{Y_{\text{pre}}}$ do $\hat{\text{var}}_{\text{całkowita}}$ i przemnożyć przez 100%:

$$\frac{\hat{\text{var}}_{Y_{\text{pre}}}}{\hat{\text{var}}_{\text{całkowita}}} \times 100\%$$

Podobnie postępujemy z pozostałymi wariancjami składowymi.

Sumując, stwierdzamy, że pretest Y zadziałał uwrażliwiająco na osoby badane, jeżeli wartość statystyki testu F -ANOVA jest większa (lub równa) od wartości krytycznej F na danym poziomie istotności statystycznej „alfa” (np. 0,05). Natomiast o wielkości tego efektu orzekamy na podstawie oszacowanych, procentowo wyrażonych, wariancji składowych: (b) – (c).

MODEL WIELOKROTNEJ REGRESJI LINIOWEJ A KONTROLA PRETESTU Y
W PLANIE SOLOMONA

Do takich samych rezultatów badawczych, jak opisane w poprzednim podpunkcie może badacz dojść stosując model wielokrotnej korelacji – regresji (MCR)². Ponieważ obie zmienne są jakościowe (dokładniej: dychotmiczne), więc przed ich wprowadzeniem do modelu niezbędne jest przeprowadzenie zabiegu kodowania. Utworzone w ten sposób nowe zmienne są zmiennymi instrumentalnymi. Umożliwiają one wykorzystanie w takich przypadkach modelu korelacyjno-regresyjnego. Z trzech znanych systemów kodowania zmiennych jakościowych (i ich interakcji) opisanych np. przez Pedhazura (1982, rozdz. 9 i 10, s. 271 i nast.) wybrałem system: „effect coding” (z uwagi na to, że system ten nawiązuje do koncepcji efektu eksperymentalnego w planach ANOVA)³.

Równanie regresji liniowej (model nieaddytywny), które badacz musi w tym przypadku zbudować przyjmuje postać:

$$Y'_{\text{post}} = b_{Y1.23}X + b_{Y2.13}Y_{\text{pre}} + b_{Y3.12}Y_{\text{pre}}X + a$$

² Podstawowe informacje na temat modelu wielokrotnej korelacji-regresji podaje m.in. Blalock (1975, rozdz. 19). W skondensowanej postaci podstawowe założenia, wzory, procedury obliczeniowe oraz odmiany MCR omówione są w: Brzeziński (1985).

³ Jeżeli mamy dostęp do EMC , to bardzo znana biblioteka programów komputerowych, adresowana do przedstawicieli nauk społecznych $SPSS$ (Nie *et al.* 1975) zawiera program na wprowadzenie zmiennych jakościowych do równania regresji wielokrotnej.

Po zakodowaniu zmiennych, które polega tu na przypisaniu danej kombinacji poziomów zmiennych X i Y_{pre} wagi liczbowej: „1” lub „-1” otrzymujemy nową tabelę wyników, taką jak przedstawiono w tabeli 3 (dla ilustracji przyjęto tu, że: $n_1 = \dots = n_4 = n = 3$)⁴. Po przeprowadzeniu obli-

Tabela 3

Tabela wyników przygotowana do przeprowadzenia analizy regresji liniowej – dane z eksperymentu przeprowadzonego wg planu Solomona ($n = 3$)

(a)	(b)	(c)	(d)	(e)	(f)
Grupy	L.p.	Y_{post}	(1)	(2)	(3) 1 x 2
XY_{pre}	1		1	1	1
	2		1	1	1
	3		1	1	1
$X \sim Y_{pre}$	4		1	-1	-1
	5		1	-1	-1
	6		1	-1	-1
$\sim XY_{pre}$	7		-1	1	-1
	8		-1	1	-1
	9		-1	1	-1
$\sim X \sim Y_{pre}$	10		-1	-1	1
	11		-1	-1	1
	12		-1	-1	1

Oznaczenia: wektor Y_{post} – wyniki posttestu Y
 wektor 1 – zakodowana zmienna X
 wektor 2 – zakodowana zmienna Y_{pre}
 wektor 3 – zakodowana interakcja zmiennych: X i Y_{pre}

czeń wg metody opisanej w cytowanej pracy Pedhazura (por. też przyp. 3) otrzymujemy tabelę sumaryczną analizy wariancji, w której (w kolumnie b) zawarte są procentowo wyrażone poszczególne wielkości wariancji składowych (ich wskaźnikami są tu współczynniki determinacji wielokrotnej – $R^2_{Y.1,\dots,n}$).

⁴ Wektor (1) utworzono w ten sposób, że każdej osobie z grupy z postępowaniem eksperymentalnym (X) przypisano wagę „1”, a osobie z grupy kontrolnej ($\sim X$) – wagę „-1”. Wektor (2) uzyskano przez przypisanie osobom o wagach „1” w wektorze (1) wagi „1”, jeżeli należą one równocześnie do grupy z pretestem Y (Y_{pre}) i wagi „-1”, jeżeli należą one równocześnie do grupy bez pretestu ($\sim Y_{pre}$). Analogicznie, osobom o wagach „-1” w wektorze (1) przypisano wagi „1”, jeżeli należą one równocześnie do grupy z pretestem (Y_{pre}) i wagi „-1”, jeżeli należały one równocześnie do grupy bez pretestu ($\sim Y_{pre}$). Wektor (3) uzyskano mnożąc przez siebie, kolejno dla każdej osoby, wartości wektorów: (1) i (2), np. dla osoby 10 mamy: $(-1) \cdot (-1) = 1$.

Tabela 4

Tabela ANOVA dla rezultatów analizy regresji liniowej (do danych z tabeli 3)

(a)	(b)	(c)	(d)	(e)	(f)
źródło wariancji	% var Y_{post}	SS	df	MS	F
wektor 1 (zmienna X)	$(R^2_{Y_{\text{post}.1}}) 100\%$				
wektor 2 (zmienna Y_{pre})	$(R^2_{Y_{\text{post}.2})} 100\%$				
wektor 3 (interakcja: XY_{pre})	$(R^2_{Y_{\text{post}.3})} 100\%$				
reszta	$1 - (R^2_{Y_{\text{post}.123})} 100\%$				
całkowita	100 %				

Za pomocą współczynnika determinacji wielokrotnej można wyznaczyć procent wariancji zmiennej zależnej (posttest Y) wyjaśniony przez poszczególne zmienne niezależne i ich interakcję. Za pomocą testu F -ANOVA oceniamy istotność zmiennych i ich interakcji dla Y_{post} . Rozkład wariancji składowych przedstawia się następująco:

$$\begin{array}{lcl}
 \text{var}(Y_{\text{post}} \mid X) & = & (R^2_{Y_{\text{post}.1})100\% \\
 \text{var}(Y_{\text{post}} \mid Y_{\text{pre}}) & = & (R^2_{Y_{\text{post}.2})100\% \\
 \text{var}(Y_{\text{post}} \mid Y_{\text{pre}}X) & = & (R^2_{Y_{\text{post}.3})100\% \\
 \text{var}(Y_{\text{post}} \mid \text{reszta}) & = & (1 - R^2_{Y_{\text{post}.123})100\%
 \end{array}$$

$$\text{var } Y_{\text{post}} \quad = \quad (R^2_{Y_{\text{post}.123})100\%$$

MODEL ANALIZY KOWARIANCJI (ANCOVA), A KONTROLA PRETESTU Y
W DWUGRUPOWYM PLANIE EKSPERYMENTALNYM

Kończąc rozważania na temat stopnia wpływu (interferencyjnego) pretestu Y na wyniki posttestu Y oraz różnorodnych prób jego określenia i oddzielenia od wpływu postępowania eksperymentalnego na wyniki posttestu Y rozważmy jeszcze jedną sytuację – nie tak znowu rzadką. Otóż co zrobić, gdy koszt badania jednej osoby jest wysoki i badacz nie dysponuje takimi środkami, które pozwalałyby mu na „rozrzutne” (w tej konkretnej sytuacji!) przebadanie, aż czterech grup porównawczych – tak jak tego wymaga rekomendowany tu plan Solomona? Można sięgnąć po plan dwugrupowy przedstawiony na rys. 1. Opisaną niżej procedurę statystyczną umożliwiła badaczowi oszacowanie „czystego” wpływu postępowania eksperymentalnego (X) na wyniki posttestu Y (Y_{post}) oddzielonego od interferencyjnego wpływu na Y_{post} pretestu Y . Zaletą podejścia niżej opisanego jest to, że następuje wyrównanie grup porównawczych (tu: eksperymentalnej i kontrolnej), jeżeli różnią się one pod względem średnich wyników w preteście Y .

Jest to bardzo ważna zaleta, zwłaszcza w sytuacji, gdy badacz nie sprawdził (z różnych powodów) równoważności grup pod względem wyników pretestu Y [porównanie: „ $Y_{pre(e)} \stackrel{s}{=} Y_{pre(k)}$ ”].

Analiza kowariancji (*ANCOVA*), stanowiąca połączenie analizy wariancji i analizy wielokrotnej regresji liniowej (por. Winer 1971; Cooley i Lohnes 1971; Timm 1975; Nie *et al.* 1975) jest tym właśnie zalecanym tu modelem analizy statystycznej danych. *ANCOVA* jest modelem statystycznym, który bywa zalecany do kontroli zmiennych ubocznych (w terminologii *ANCOVA*: „zmiennie towarzyszące”) w badaniach efektów nauczania (Werts, Linn, 1971), czy w badaniach nad rozwojem (Brzeziński, 1988). *ANCOVA* jest zalecana zwłaszcza wówczas, gdy współczynnik korelacji wyników pretestu Y i posttestu Y przekroczy wartość 0,60 (wg Feldt 1958). W praktyce badawczej często spotykamy się z wartościami współczynników korelacji $r_{(pre)(post)} > 0,70$. Może być ona wykorzystana z pożytkiem dla wyników badania, jeżeli będą spełnione dość restryktywne założenia, a mianowicie założenie o losowym wyznaczeniu poziomów postępowania eksperymentalnego do grup porównawczych (Randomizacja II), założenie liniowości, założenie homogeniczności współczynników regresji (inaczej: brak interakcji między pretestem Y i postępowaniem eksperymentalnym X !) oraz założenie o interwałowym poziomie pomiaru Y_{pre} i Y_{post} .

Wyniki badania umieszczamy w tabeli takiej jak Tab. 5. Następnie

Tabela 5

Tabela wyników surowych wymagana w analizie kowariancji – dane z eksperymentu przeprowadzonego wg planu dwugrupowego z rys. 1

(a)	(b)		(c)	
Osoby	X		~X	
	Y_{pre}	Y_{post}	Y_{pre}	Y_{post}
1				
2				
⋮				
⋮				
n				
suma				
średnia				

dokonyjemy zabiegu kodowania zmiennej X (zmienna: Y_{pre} jest tu *zmienną towarzyszącą* i ze względu na interwałowy poziom pomiaru, podobnie jak i zmienna Y_{post} nie wymaga przeprowadzenia zabiegu kodowania) – np. według systemu: „effect coding”. Po wykonaniu zabiegu kodowania otrzymujemy tabelę (por. tabela 6) gotową do przeprowadzenia analizy

Tabela 6

Tabela wyników zakodowanych (na podstawie tabeli 5), przygotowana do przeprowadzenia analizy kowariancji

(a)	(b)	(c)	(d)	(e)	(f)
Postępowanie eksperymentalne	L.p.	Y_{pre}	(1)	(2) $Y_{pre} \times 1$	Y_{post}
X	1		1		
	2		1		
	3		1		
	4		1		
	5		1		
~X	6		-1		
	7		-1		
	8		-1		
	9		-1		
	10		-1		

Oznaczenia: wektor Y_{post} – wyniki posttestu Y
 wektor Y_{pre} – wyniki pretestu Y (zmienna towarzysząca)
 wektor 1 – zakodowana zmienna X
 wektor 2 – zakodowana interakcja zmiennych X i Y_{pre}

kowariancji (wg Pedhazur 1982). Rezultaty przeprowadzonej analizy kowariancji zamieszczamy w sumarycznej tabeli (por. tabela 7). Kolumna „b” zawiera procentowe wartości poszczególnych wariacji składo-

Tabela 7

Tabela ANCOVA (dla danych z tabeli 5)

(a)	(b)	(c)	(d)	(e)	(f)
źródło wariacji	% var Y_{post}	SS	df	MS	F
zmienna towarzysząca (Y_{pre})	$(R^2_{Y_{post} \cdot Y_{pre}})100\%$				
postępowanie eksperymentalne – adj. (X – adj.)	$(R^2_{Y_{post} \cdot Y_{pre} \cdot 1} - R^2_{Y_{post} \cdot Y_{pre}})100\%$				
reszta	$(1 - R^2_{Y_{post} \cdot Y_{pre}})100\%$				
całkowita	100%				

wych. Ostatnia kolumna informuje badacza o istotności (via test F) poszczególnych zmiennych (symbol: „adj.” od angielskiego – *adjusted* – informuje, iż podane w tym wierszu wartości: „% var Y_{post} ”, „SS”, „F” zostały skorygowane poprzez wyeliminowanie wpływu zmiennej towarzy-

szącej na wyniki posttestu Y). Badacz może porównać jeszcze średnie wyniki posttestu Y , nie skorygowane, ze średnimi wynikami posttestu Y , skorygowanymi (*adj.*). Porównanie to informuje badacza o stopniu wyrównania grup, eksperymentalnej i kontrolnej pod względem wyników pretestu Y .

Podstawowa zaleta ANCOVA, tj. skorygowanie różnic między średnimi wynikami posttestu Y , z uwagi na występujące różnice między średnimi wynikami pretestu Y nie powinna przysłańcać badaczowi (zwłaszcza zwolennikowi tego podejścia) faktu, iż oszczędność środków uzyskana dzięki redukcji liczby grup, z czterech – plan Solomona – do dwóch, pociąga za sobą niemożność oddzielenia wpływu pretestu Y na przygotowanie osób badanych na „przyjęcie” postępowania eksperymentalnego, a tym samym na kierunek i wielkość zmiany. Pozostaje tedy nadal otwartym pytanie: co w wynikach posttestu Y da się „wytłumaczyć” zmienną Y_{pre} , co interakcją XY_{pre} , a co – zmienną X ? Tak więc idealnym rozwiązaniem pozostaje przeprowadzenie badania wedle planu Solomona.

5. PODSUMOWANIE

W niniejszym artykule przedyskutowano problemy metodologiczne jakie rodzi stosowanie pretestu zmiennej Y w eksperymentalnym badaniu zmian. Doprowadza nas to do następujących wniosków.

1. Do oceny stopnia zmiany, jaka zaszła pod wpływem postępowania eksperymentalnego, badacz nie powinien poprzestać tylko na porównaniu grupy eksperymentalnej z kontrolą pod względem średnich wyników posttestu zmiennej zależnej. Ważne bowiem jest udzielenie odpowiedzi na pytanie: jak wielka zaszła zmiana między pierwszym i drugim pomiarem zmiennej zależnej w grupie eksperymentalnej i w jakiej ona pozostaje relacji do zmiany zaobserwowanej w grupie kontrolnej?

2. Pretest zmiennej Y może być źródłem artefaktów w badaniach społecznych, zwłaszcza wówczas, gdy badacz posługuje się takimi narzędziami pomiaru zmiennej Y , które mają właściwości uwrażliwiające osoby badane (np. testy inteligencji, skale postaw itp.). Jest on czynnikiem zakłócającym trafność wewnętrzną oraz trafność zewnętrzną planu eksperymentalnego.

3. Rozpatrzono różne możliwe porównania, w których występują wyniki pretestu Y , a które mogą być wykorzystane przez badacza jako wskaźniki zmiany (bądź indywidualne, bądź – via średnie arytmetyczne – grupowe). W szczególności zwrócono uwagę na wyniki złożone, takie jak: wyniki różnicowe i wyniki resztowe. Cechuje je stosunkowo niska rzetelność i szczególne uzależnienie od stopnia korelacji wyników pretestu Y i posttestu Y .

4. Rozwiązanie problemu pretestu przez przeprowadzenie badań wg

planów eliminujących pretest Y nie jest dobrym rozwiązaniem. Należy przeprowadzić pretest Y i poddać go kontroli badacza w planie badawczym.

5. Najodpowiedniejszym (ale i kosztownym) schematem badań eksperymentalnym umożliwiającym pełną kontrolę efektu pretestu Y jest czterogrupowy plan Solomona.

6. Zaleca się dwie równoważne metody statystycznego opracowania danych uzyskanych z badania przeprowadzonego wedle planu Solomona: (a) zastosowanie dwuczynnikowej analizy wariancji połączonej z analizą wariancji składowych oraz (b) zastosowanie wielokrotnej regresji liniowej (model nieaddytywny).

7. Jeżeli badacz nie wyrównał grup porównawczych pod względem wyników pretestu Y , to może on nadal prowadzić badanie, ale jego wyniki powinien opracować za pomocą analizy kowariancji. Pretest Y jest tu traktowany jako *zmienna towarzysząca*. Otrzymana procentowa wartość wariancji składowej tłumaczona działaniem postępowania eksperymentalnego na Y „oczyszczona” została z interferencyjnych wpływów pretestu Y na wyniki posttestu Y . Ponieważ jednak pretest Y przeprowadzono w obu grupach (czy większej ich liczbie) więc nie można udzielić odpowiedzi na pytanie: czy grupa bez pretestu Y otrzymałaby takie same wyniki? Tak więc analiza kowariancji zapewnia tylko kontrolę wyrównania grup porównawczych pod względem wyników pretestu Y .

8. Wspomniano też, chociaż myśl ta z uwagi na ramy artykułu nie była rozwijana, że zwiększyć można efektywność badań eksperymentalnych nad zmianą przez zwiększenie liczby posttestów Y (badanie dynamiki zmiany!). Modele analizy wariancji, analizy kowariancji i analizy regresji wielokrotnej w wersji z powtarzаныmi pomiarami zmiennej zależnej znajdują tu zastosowanie (por. Brzeziński, 1988; Cohen, Cohen 1975, rozdz. 10; Bock 1963, 1975; Timm 1975, 1980; Hedayat, Afsarinejad 1975).

czerwiec, 1985

LITERATURA

- Babad, E. Y., Inbar, J., Rosenthal R., (1982), *Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers*. „Journal of Educational Psychology”, 74, s. 459 – 474.
- Batles, P. B. (red.), (1978), *Life-span development and behavior* t. 1, New York.
- Batles, P. B., Brim, O. G., Jr. (red.), (1979), *Life-span development and behavior*, t. 2, New York.
- Batles, P. B., Brim O. G., Jr. (red.), (1980), *Life-span development and behavior*, t. 3, New York.
- Batles, P. B., Brim. O. G., Jr. (red.), (1981), *Life-span development and behavior*, t. 4, New York.
- Batles, P. B., Brim, O. G., Jr. (red.), (1983), *Life-span development and behavior*, t. 5, New York.
- Bereiter, C. (1963), *Some persisting dilemmas in the measurement of change*, [w:] C. W. Harris (red.), *Problems in measuring change*, s. 3 – 20, Madison.
- Blałock, H. M., (1975), *Statystyka dla socjologów*, Warszawa.
- Bock, R. D. (1963), *Multivariate analysis of variance of repeated measurements*, [w:] C.W. Harris (red.), *Problems in measuring change*, s. 85 – 103, Madison.
- Bock, R. D., (1975), *Multivariate statistical methods in behavioral research*, New York.
- Brzeziński, J., (1975), *Eksperymentalna kontrola skuteczności zabiegów psychokorekcyjnych*, „Przegląd Psychologiczny”, 18, s. 557 – 565.
- Brzeziński, J., (1978), *Metodologiczne i psychologiczne wyznaczniki procesu badawczego w psychologii*, Poznań.
- Brzeziński, J., (1984), *Elementy metodologii badań psychologicznych*, Wyd. 4, Warszawa.
- Brzeziński, J., (1983), *Ocena efektu eksperymetalnego w układach eksperymetalnych analizy wariancji*, „Przegląd Psychologiczny”, 25, s. 155 – 165.
- Brzeziński, J., (1975), *Zasady statystycznego opracowania danych w psychologii klinicznej*, [w:] B. Waligóra (red.), *Wybrane zagadnienia z psychologii klinicznej*, s. 107 – 160, Poznań.
- Brzeziński, J. (w druku – a), *Statystyczna kontrola „efektu pretestu” w badaniu zmian, „Człowiek i Społeczeństwo”*.
- Brzeziński, J. (1985), *Zastosowanie liniowego modelu korelacyjno-regresyjnego w badaniach psychologicznych (wybrane zagadnienia)*, [w:] E. Paszkiewicz, T. Szustrowa (red.), *Materiały do nauczania psychologii*, seria III, t. 4, Warszawa.
- Brzeziński, J. (1988), *Statystyczne modele badania zmian rozwojowych (analiza wariancji, analiza kowariancji, analiza regresji wielokrotnej)*, [w:] M. Tyszkowa (red.), *Rzecz psychiczny w ciągu życia. Zagadnienia teoretyczne i metodologiczne*, Warszawa.
- Brzeziński, J., Maruszeński, T., (1981), *Nieparametryczne analizy statystyczne w protoidealizacyjnym modelu nauki*, „Kwartalnik Pedagogiczny”, 1, s. 59 – 75.
- Brzeziński, J., Stachowski, R., (1984), *Zastosowanie analizy wariancji w eksperymetalnych badaniach psychologicznych*, Warszawa.
- Campbell, D. T. (1969), *Reforms as experiments*, „American Psychologist”, 24, s. 409 – 429.
- Campbell, D. T., Stanley, J. C., (1967), *Experimental and quasi-experimental designs for research on teaching*, [W:] N. L. Gage (red.), *Handbook of research on teaching*, s. 171 – 246, Chicago.
- Choynowski, M., (1971), *Podstawy i zastosowania teorii rzetelności testów psychologicznych*, [w:] J. Koziński (red.), *Problemy psychologii matematycznej*, s. 65 – 118, Warszawa.
- Cohen, J., (1983), *The cost of dichotomization*, „Applied Psychological Measurement”, 7, s. 249 – 253.

- Cohen, J., Cohen, P., (1975), *Applied multiple regression/correlation analysis for the behavioral sciences*, Hillsdale, N.J.
- Cooley, W., Lohnes, P., (1971), *Multivariate procedures for the behavioral sciences*, New York.
- Czabała, C., Leder, S., Pohorecka, A., (1973), *Badania nad psychoterapią grupową*, [w:] H. Wardaszko-Łyskowska (red.), *Terapia grupowa w psychiatrii*, s. 181 – 209, Warszawa.
- Domański, C., (1979), *Statystyczne testy nieparametryczne*, Warszawa.
- Draper, N. S., Smith, H., (1973), *Analiza regresji stosowana*, Warszawa.
- Endler, N. S., (1966), *Estimating variance components from mean squares for random and mixed effects analysis of variance models*, „Perceptual and Motor Skills”, 22, s. 559 – 570.
- Feldt, L. S., (1958), *A comparison of the precision of three experimental designs employing a concomitant variable*, „Psychometrika”, 23, s. 335 – 353.
- Fisher, R. A., (1925), *Statistical methods for research workers*, London.
- Fisher, R. A., (1935), *The design of experiments*, London.
- Geras, G., (1985), *Elektrodermometria w psychologii i medycynie*, Gdańsk.
- Crump, S. L., (1951), *The present status of variance components analysis*, „Biometrics”, 7, s. 1 – 16.
- Gulliksen, H., (1950), *Theory of mental tests*, New York.
- Halderson, J. S., Glasnapp, D. R. (1972), *Generalized rules for calculating the magnitude of an effect in factorial and repeated measures ANOVA designs*. „American Educational Research Journal”, 9, s. 301 – 310.
- Hays, W. L., (1973), *Statistics for the social sciences*, New York.
- Hedayat, A., Afsarinejad, K., (1975), *Repeated measurements design*, [w:] J. N. Srivastava (red.), *A survey of statistical designs*, s. 229 – 242, Amsterdam.
- Hollander, H., Wolfe, D. A., (1973), *Nonparametric statistical methods*, New York.
- Krajewski, Wł., (1977), *Pojęcia rozwoju i postępu*, [w:] J. Kmita (red.), *Założenia teoretyczne badań nad rozwojem historycznym*, s. 21 – 45, Warszawa.
- Lehmann, E. L., (1968), *Testowanie hipotez statystycznych*, Warszawa.
- Linn, R. L., Slind, J. A. (1977), *The determination of the significance of change between pre- and posttesting periods*, „Review of Educational Research”, 47, s. 121 – 150.
- Lord, F. M., (1963), *Elementary models for measuring change*, [w:] C. W. Harris (red.), *Problems in measuring change*, s. 21 – 38, Madison.
- Lord, F. M., Novick, M. R., (1968), *Statistical theories of mental test scores*, Reading, Mass.
- Muszyński, H. (1971), *Wstęp do metodologii pedagogiki*, Warszawa.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., Bent, D. H. (1975). *SPSS. Statistical Package for the Social Sciences*, New York.
- O'Connor, E. F., Jr., (1972), *Extending classical test theory to the measurement of change*, „Review of Educational Research”, 42, s. 73–97.
- Oktała, W., (1966), *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, Warszawa.
- Orne, M. T., (1962), *On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications*. „American Psychologist”, 17, s. 776–783.
- Orne, M. T., (1969), *Demand characteristics and the concept of quasi-controls*, [w:] R. Rosenthal, R. Rosnow (red.), *Artifact in behavioral research*, s. 143–179, New York.
- Orne, M. T., (1970), *Hypnosis, motivation, and the ecological validity of the psychological experiment*, [w:] W. J. Arnold, M. M. Page (red.), *Nebraska Symposium on motivation*, s. 187–265, Lincoln.
- Ossowski, St. (1967), *O osobliwościach nauk społecznych*, [w:] St. Ossowski, *Dzieła t. 4*, s. 125–316, Warszawa.

- Owerall, J. E., Woodward, J.A., (1975), *Unreliability of difference scores: A paradox for measurement of change*, „Psychological Bulletin”, 82, s. 85–86.
- Pedhazur, E. J., (1982), *Multiple regression in behavioral research. Explanation and prediction*, wyd. 2, New York.
- Rosenberg, M. J., (1965), *When dissonance fails: On eliminating evaluation apprehension form attitude measurement*, „Journal of Personality and Social Psychology”, 1, s. 28–42.
- Rosenberg, M. J., (1969), *The conditions and consequence of evaluation apprehension*, [w:] R. Rosenthal, R. Rosnow (red.), *Artifact in behavioral research*, s. 279–349, New York.
- Rosenthal, R., (1969), *Interpersonal expectations: Effects of the experimenter's hypothesis*, [w:] R. Rosenthal, R. Rosnow (red.), *Artifact in behavioral research*, s. 181–277, New York.
- Rosenthal, R., (1976), *Experimenter effects in behavioral research*, wyd. zm., New York.
- Rosenthal, R., Rubin, D. B. (1978), *Interpersonal expectancy effects: the first 345 studies*, „The Behavioural and Brain Sciences”, 3, s. 377–415.
- Sechrest, L., Yeaton, W. H., (1982), *Magnitudes of experimental effects in social science research*, „Evaluation Review”, 6, s. 579–600.
- Siegel, S., (1956), *Nonparametric statistics for the behavioral sciences*, New York.
- Solomon, R., (1949), *An extension of control group design*, „Psychological Bulletin”, 46, s. 137–150.
- Stevens, S. S., (1951), *Mathematics, measurement and psychophysics*, [w:] S. S. Stevens (red.), *Handbook of experimental psychology*, s. 1–49, New York.
- Thorndike, R. L., (1966), *Intellectual status and intellectual growth*, „Journal of Educational Psychology”, 57, s. 121–127.
- Timm, N. H., (1975), *Multivariate analysis with applications in education and psychology*, Monterey, CA.
- Timm, N. H., (1980), *Multivariate analysis of variance of repeated measurements*, [w:] P. R. Krishnaiah (red.), *Handbook of statistics*, t. 1, s. 41–87, Amsterdam.
- Werts, C. E., Linn, R. L., (1971), *Analyzing school effects: ANCOVA with a fallible covariate*, „Educational and Psychological Measurement”, 31 s. 95–104.
- Winer, B. J., (1971), *Statistical principles in experimental design*, wyd. 2, New York.