

## **ROZDZIAŁ 12**

### **SCHEMATY BADAŃ – METODY EKSPERYMENTALNE I QUASI-EKSPERYMENTALNE**

#### **1. Wstęp**

Jak już wcześniej była mowa, jednym z podstawowych celów ewaluacji działań podejmowanych przez służby społeczne jest sprawdzenie, czy działania te przynoszą zamierzony skutek lub szerzej jakie wywołują efekty zamierzone i niezamierzone. Jest to kluczowe z punktu widzenia oceny tych działań jako mających sens, zasadnych czy potrzebnych ludziom, do których są skierowane. Gdy podejmowane są takie działania jak np. pomoc osobom bezrobotnym w poszukiwaniu pracy, zajęcia terapeutyczne dla ofiar przemocy w rodzinie, pomoc osobom uzależnionym w wyrwaniu się z nałogu, pomoc osobom bezdomnym w wyjściu z bezdomności czy warsztaty edukacyjne dla dzieci mających problemy w nauce, należy stwierdzić czy rzeczywiście działania te służą realizacji celów, dla których są podejmowane. Działania te z kolei mogą być prowadzone w różny sposób, różnymi metodami. Istotne jest więc określenie czy zastosowana metoda lub metody są skuteczne i która ze stosowanych metod jest najskuteczniejsza. Aby móc rzetelnie określić skuteczność lub szerzej rezultat określonego działania (wybranej metody działania), nie wystarczy jedynie obserwować sytuację sprzed i po wykonaniu danej interwencji<sup>1</sup>. W ten sposób nie da się bowiem stwierdzić, czy na pewno zmiana która nastąpiła (jeśli nastąpiła) była wynikiem tego działania. Innymi słowy, na tej podstawie nie da się stwierdzić występowania związku przyczynowo-skutkowego między danym działaniem a obserwowaną zmianą. Możemy jedynie ustalić współwystępowanie danego działania i zmiany. Wynika to z kilku względów (tzw. zakłóceń oddziaływania bodźca).

- Po pierwsze, zaobserwowana zmiana może być dziełem przypadku<sup>2</sup>.
- Po drugie, może być ona również wynikiem wystąpienia (oddziaływania) innych czynników, niż podejmowane działanie<sup>3</sup>.
- Po trzecie, w obserwowanej rzeczywistości mogą zachodzić systematyczne zmiany, które dokonują się w niej w miarę upływu czasu, bez wpływu konkretnych wydarzeń (czyli podlegają pewnej długofalowej tendencji, pewnemu trendowi).

---

<sup>1</sup> Taka obserwacja (pomiar) sytuacji sprzed i po przeprowadzeniu jakiegoś działania (czyli wprowadzeniu bodźca) nazywana jest w socjologii odpowiednio: pretestem i posttestem.

<sup>2</sup> Czyli tzw. niekontrolowanych zdarzeń losowych.

<sup>3</sup> Są to tzw. zmiany zewnętrzne.

- Po czwarte, może zdarzyć się, że początkowo obserwowana sytuacja była skrajna (np. szczególnie zła), czyli wystąpiło tzw. chwilowe odchylenie od trendu, które powróciło na ścieżkę trendu w trakcie lub po wprowadzeniu działania, lecz bez związku z nim (jest to tak zwany efekt regresji).
- Po piąte, zmianie może ulec sposób dokonywania obserwacji (w szczególności gdy zmieniają się zasady lub narzędzia pomiaru) sytuacji sprzed i po wprowadzeniu działania, co wpływa na wystąpienie różnicy w wynikach obserwacji, a nie jest przecież rezultatem tego działania.

W ramce nr 1 zaprezentowano przykład obrazujący wyżej wymienione trudności w określaniu skuteczności danego działania.

#### **Ramka 1**

##### **Przykład 1. Trudności w określeniu wpływu programu edukacyjnego na wyniki w nauce dzieci objętych tym programem.**

W pewnym Ośrodku Pomocy Społecznej wprowadzono nowatorski program pomocy dzieciom mającym problemy w nauce i pochodzącym z ubogich rodzin, dofinansowany ze środków Europejskiego Funduszu Społecznego. Program miał postać pięciogodzinnych warsztatów edukacyjnych odbywających się przez jeden semestr (letni), w każdą sobotę, pod kierunkiem wykwalifikowanych nauczycieli oraz psychologów. Jednak ze względu na ograniczoność środków finansowych programem objęto tylko część (ośmioro) dzieci mających problemy w nauce z rodzin korzystających z pomocy tego ośrodka. Podstawowym celem programu było doprowadzenie do sytuacji, w której dzieci objęte tym programem będą lepiej radziły sobie w nauce. Za wskaźnik "radzenia sobie w nauce" przyjęto średnią ocen. Po semestrze trwania programu porównano wyniki w nauce dzieci biorących w nim udział z tymi z końca poprzedniego semestru. Okazało się, że czwórka dzieci uzyskała lepsze wyniki, trójka dzieci miała wyniki podobne do tych sprzed roku a jedno dziecko nieco obniżyło się w nauce. Średnia ocen dla całej grupy podniosła się nieco (o 0,375 stopnia).

Czy na podstawie takiego porównania możemy jednoznacznie stwierdzić, że polepszenie wyników części dzieci było rezultatem oddziaływania programu? Niestety nie jest to takie oczywiste. Po pierwsze to, że oceny części dzieci uległy poprawie, może być dziełem przypadku. Po drugie, lepsze oceny części dzieci mogą być wynikiem wpływu innych czynników niż program, np. w szkole do której uczęszczają dzieci zmienił się nauczyciel od nauki polskiego oraz wychowawca, a materialna sytuacja rodzinna dwójki dzieci uległa poprawie, dzięki czemu rodziców stać było na zakupienie dzieciom pomocy szkolnych (m.in. komputera), których wcześniej nie posiadały. Po trzecie, nie można wykluczyć, że poprawa wyników części dzieci była wynikiem pozytywnej tendencji (np. dojrzewania dzieci), a także tego, że przed rozpoczęciem udziału w programie sytuacja dzieci była wyjątkowo zła (skrajne odchylenie od trendu) i po kilku miesiącach wróciła do normy. Ostatecznie trzeba również uwzględnić fakt, że wynik pomiaru średniej ocen może nie być porównywalny ze względu na zmianę sposobu liczenia średniej (w drugim semestrze, czyli na koniec roku szkolnego, wliczono do średniej ocen również ocenę z religii).

Źródło: opracowanie własne.

Naukową odpowiedzią na wyżej opisane trudności w określaniu wpływu danego działania na grupę osób (sytuację) jest wykonanie eksperymentu. Eksperyment, to "powtarzalny zabieg, polegający na planowej zmianie przez badacza jednych czynników w badanej sytuacji, przy równoczesnej kontroli innych czynników, podjęty w celu uzyskania w drodze obserwacji odpowiedzi na pytanie o skutki tej zmiany"<sup>4</sup>. Najważniejszym celem eksperymentu jest sprawdzenie hipotez o związkach przyczynowych, czyli np. hipotezy o tym, że opisany w przykładzie 1 program edukacyjny ma pozytywny wpływ na wyniki szkolne dzieci objętych tym programem.

Poniżej zaprezentowano podstawowe zasady przeprowadzania klasycznego eksperymentu stochastycznego oraz wybranych jego modyfikacji wraz z opisem tego, jak mógłby wyglądać eksperyment przeprowadzony w celu określenia wpływu opisanego w przykładzie nr 1 programu edukacyjnego na wyniki w nauce dzieci objętych tym programem.

## 2. Klasyczny eksperyment stochastyczny

Klasyczny schemat eksperymentalny w swej najprostszej postaci wymaga skonstruowania dwóch grup porównawczych - **grupy eksperymentalnej i grupy kontrolnej**, ujednoliconych pod istotnymi względami, zwłaszcza pod względem tzw. zmiennej zależnej, a więc pod względem tego, na co ma mieć wpływ badane przez nas działanie (bodziec). Grupy te mogą składać się z osób, zbiorowości lub instytucji - przy czym w pracy służb społecznych będą to przede wszystkim ludzie lub zbiorowości. **Ujednolicenie grup w schemacie klasycznym** można uzyskać stosując następujące metody, choć przez część badaczy jedynie pierwsza uznawana jest za schemat w pełni eksperymentalny<sup>5</sup>:

- randomizację, czyli losowy<sup>6</sup> podział jednostek badania na grupę eksperymentalną i kontrolną;
- dobór parami, czyli grupowanie jednostek badania w pary o jednakowych wartościach istotnych zmiennych (cech) i następnie losowe rozdzielenie tych par na jednostki eksperymentalne i kontrolne;
- wyrównywanie częstości względnych, tj. celowy dobór grup o jednakowych rozkładach ważnych zmiennych<sup>7</sup> i wylosowanie jednej z tych grup jako eksperymentalnej, a drugiej jako kontrolnej.

---

<sup>4</sup> Sulek A., *Badania eksperymentalne i quasi-eksperymentalne*, [w:] *Metody analizy socjologicznej*, Instytut Socjologii UW, Warszawa 1986, s. 57.

<sup>5</sup> por. <http://socialresearchmethods.net>.

<sup>6</sup> Losowość oznacza, że każda osoba ma takie samo prawdopodobieństwo znalezienia się w którejś z grup. Grupy dobrane losowo nie będą identyczne, ale można określić prawdopodobieństwo z którym zaobserwowane różnice w wynikach nie są wynikiem pierwotnych różnic w grupach.

- dobór grup równoważnych, czyli celowy wybór spośród danych zbiorowości dwóch zbiorowości najbardziej do siebie podobnych.

Gdy mamy już wybrane dwie grupy, eksperymentalną i kontrolną, w obu grupach dokonujemy pomiaru przynajmniej wartości zmiennej zależnej (czyli **wykonujemy tzw. pretest**). Następnie grupę eksperymentalną poddajemy oddziaływaniu bodźca, czyli danego działania z zakresu pracy służb społecznych. Po odpowiednim czasie ponownie dokonujemy pomiaru wartości zmiennej zależnej (czyli **wykonujemy tzw. posttest**). Należy oczywiście pamiętać o zastosowaniu jednolitych narzędzi pomiarowych dokonując pretest i posttest. Jeśli pretest w grupie eksperymentalnej oznaczymy symbolem  $P_{E1}$ , zaś w grupie kontrolnej  $P_{K1}$ , natomiast posttest w grupie eksperymentalnej symbolem  $P_{E2}$ , zaś w grupie kontrolnej  $P_{K2}$ , wówczas wpływ bodźca (działania) wyraża różnica ( $d$ ), którą obliczamy korzystając z poniższego wzoru:

$$d = (P_{E2} - P_{E1}) - (P_{K2} - P_{K1}).$$

Eksperymentalny schemat klasyczny radzi sobie ze wszystkimi wyżej omawianymi trudnościami w określeniu wpływu działania na badaną grupę jednostek (sytuację). Wstępne ujednoczenie grupy eksperymentalnej i kontrolnej pozwala kontrolować pozostałe (inne niż wprowadzony bodziec) czynniki uboczne oraz wewnętrzne zmiany zachodzące w grupach, ponieważ działają one na obie grupy jednakowo. Warto dodać, że w schemacie klasycznym efektywna kontrola innych czynników niż bodziec, nie wymaga uprzedniego przewidzenia ich pojawienia się - schemat ten bowiem dobrze kontroluje zarówno znane zakłócenia, jak i te których istnienia nawet się nie domyślamy. To natomiast, czy zaobserwowane różnice mogą być wytłumaczone niekontrolowanymi okolicznościami przypadkowymi, sprawdzić można wykonując odpowiedni tzw. statystyczny test istotności. Można go wykonać samodzielnie, lecz jest to dosyć skomplikowane i potrzebna jest do tego odpowiednia wiedza statystyczna<sup>8</sup>. Obecnie jednak, we wszystkich wykorzystywanych w instytucjach publicznych pakietach statystycznych (takich jak np. SPSS, Statistica, Statgraphics, SAS czy STATA) można taki test wykonać automatycznie, ważne jest jedynie by go odpowiednio zinterpretować. Poniżej

---

<sup>7</sup> Rozkład zmiennej to sposób w jaki rozkładają się wartości danej cechy pośród jednostek zbiorowości. Np. jeśli chcemy porównać dwie grupy dzieci pod względem wyników w nauce, to jedną z ważnych cech tych dzieci będzie oczywiście wiek. Wyrównując częstości względne w grupach, które tworzymy dobieramy do grup dzieci w taki sposób, aby w każdej grupie były takie same odsetki dzieci w poszczególnych przedziałach wiekowych.

<sup>8</sup> Zainteresowanych odesłać można do następujących podręczników statystycznych: J. Podgórski, J. Józwiak, *Statystyka od podstaw*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 2009; M. Rószkiewicz, *Statystyka. Kurs podstawowy*, 2002.

zaprezentowano (ramka 2) przykład zastosowania klasycznego eksperymentu oraz interpretacji wyników testu statystycznego dla oceny wpływu omówionego w przykładzie 1 (ramka 1) programu edukacyjnego.

#### Ramka 2

### Przykład 2. Zastosowanie klasycznego eksperymentu do oceny wpływu programu edukacyjnego na wyniki w nauce dzieci objętych tym programem.

Aby rzetelnie ocenić wpływ opisanego w przykładzie 1 programu edukacyjnego na wyniki w nauce dzieci, można przeprowadzić klasyczny schemat eksperymentalny. W tym celu powinniśmy dokonać wyboru dwóch grup dzieci, grupy eksperymentalnej, która zostanie objęta programem oraz grupy kontrolnej. W związku z tym, że ograniczone środki finansowe pozwalają na objęcie programem tylko ośmiu dzieci, grupa eksperymentalna składać się będzie z takiej właśnie liczby jednostek. Podobna liczebnie powinna być również grupa kontrolna. Dokonując doboru dzieci do poszczególnych grup, wykorzystamy randomizację, czyli losowy podział dzieci na grupy, co będzie sprawiedliwe i pozwoli nam uniknąć dyskryminacji jednych dzieci względem drugich.

W praktyce losowanie możemy przeprowadzić na różne sposoby. Dla przykładu, w programie Excel tworzymy listę wszystkich dzieci objętych opieką danego OPS, które mogłyby uczestniczyć w programie. Następnie każdemu dziecku, wykorzystując w tym celu funkcję ( $f(x)$ ) LOS, nadajemy liczbę z przedziału od 0 do 1 (program dokonuje tego w sposób losowy). W końcu porządkujemy listę dzieci względem nadanych liczb - od najmniejszej do największej i do grupy eksperymentalnej przyporządkowujemy np. pierwsze osiem osób, zaś do grupy kontrolnej kolejne osiem. Kolejny krok to wykonanie pretestu w obu grupach, czyli obliczenie średniej ocen z przedmiotów po pierwszym semestrze dla każdego dziecka z obu grup oraz średniej dla każdej z grup, ze zwróceniem uwagi na to, by do średniej wchodziły te same przedmioty, które będą do niej wliczane na koniec roku. Po zakończeniu trwania programu należy ponownie dokonać pomiaru (posttest) średniej ocen i następnie porównać wyniki. Poniższa tabela przedstawia wyniki pretestu i posttestu dla obu grup:

Średnia ocen dla grup:	Pretest	Posttest
Grupa eksperymentalna	2,875	3,25
Grupa kontrolna	2,875	3,125

Miara wpływu programu na wyniki w nauce dzieci nim objętych wynosi:

$d = (3,25 - 2,875) - (3,125 - 2,875) = 0,125$  Widzimy więc, że różnica w wynikach uczniów jest rezultatem oddziaływania nie tylko badanego programu, ale w znacznej mierze również innych czynników. Zaobserwowana różnica w średniej ocen między grupami w posttestie jest mała, powinniśmy więc dodatkowo sprawdzić, czy nie jest ona dziełem przypadku. W tym celu należy przeprowadzić test statystyczny t-studenta dla grup niezależnych. Poniższa tabela prezentuje wyniki takiego testu wykonanego za pomocą pakietu statystycznego SPSS<sup>9</sup>.

<sup>9</sup> Aby dokonać tego testu, w programie SPSS 14PL utworzono bazę danych z 3 zmiennymi: zmienną "pretest" zawierającą wyniki średnich obliczonych dla wszystkich 16 dzieci (8 z grupy eksperymentalnej i 8 z kontrolnej) przed rozpoczęciem programu, zmienną "posttest" zawierającą wyniki średnich dla wszystkich 16 dzieci obliczonych po zakończeniu programu oraz zmienną "grupy" przyporządkowującą każdemu dziecku symbol "e" - jeśli należało do grupy eksperymentalnej i "k" - jeśli należało do grupy kontrolnej. Następnie wybrano z menu: Analiza -> Porównanie średnich -> Test t dla prób niezależnych. W okienku testowym wybrano zmienną "posttest" jako zmienną testowaną oraz zmienną "grupy" jako zmienną grupującą oraz określono 2 grupy: e i k.

Statystyki dla grup

grupy	N	Średnia	Odchylenie standardowe	Błąd standardowy średniej
posttest e	8	3,2500	,71962	,25443
k	8	3,1250	,23146	,08183

Test dla prób niezależnych

		Test Levene'a jednorodności wariancji		Test t równości średnich						
		F	Istotność	t	df	Istotność (dwustronna)	Różnica średnich	Błąd standardowy różnicy	95% przedział ufności dla różnicy średnich	
									Dolna granica	Górna granica
posttest	Założono równość wariancji	3,182	,096	,468	14	,647	,12500	,26726	-,44822	,69822
	Nie założono równości wariancji			,468	8,433	,652	,12500	,26726	-,48584	,73584

Okazuje się, że różnica między grupami w postteście nie jest istotna statystycznie<sup>10</sup> (ponieważ istotność jest większa niż standardowo przyjmowane  $p < 0,05$ <sup>11</sup> - formalny zapis:  $t(14)=0,468$ ;  $p < 0,96$ ), co oznacza, że nie możemy twierdzić, iż przeprowadzony program edukacyjny miał istotny wpływ na polepszenie wyników szkolnych dzieci objętych tym programem.

Źródło: opracowanie własne.

Klasyczny schemat eksperymentalny ma swoje **wady i ograniczenia**. Trudno jest nieraz spełnić jego wymagania, czyli przede wszystkim dobrać dwie ujednoczone względem siebie (jednakowe) grupy, zapewnić im jednolite warunki i zmierzyć je w dwóch, rozdzielonych wprowadzeniem bodźca momentach. Ponadto trafność eksperymentu klasycznego obniżyć może interakcja bodźca z pretestem, w sytuacji w której dokonanie pomiaru wymaga zaangażowania w niego osób badanych. Pod wpływem pretestu badani mogą uwarżliwić się na bodziec eksperymentalny i zmienić pod wpływem samego pretestu swoje zachowanie czy postawy. Co więcej, często nie da się zabezpieczyć przed pośrednim oddziaływaniem wprowadzonego w grupie eksperymentalnej bodźca na grupę kontrolną (poprzez interakcje osób należących do obu grup). Dla przykładu, dzieci biorące udział w programie edukacyjnym z przykładu nr 2 wchodziły w interakcje z dziećmi z grupy kontrolnej przekazując im część zdobytej wiedzy i umiejętności lub chociażby wywołując w nich odruch współzawodnictwa, co może być przyczyną tego, że wyniki szkolne w obu grupach

<sup>10</sup> Należy dodać, że gdyby różnica okazała się statystycznie istotna, należałoby jeszcze sprawdzić, czy różnica między pretestem i posttestem w obu grupach jest istotna statystycznie, w tym celu należałoby przeprowadzić test t-studenta dla grup zależnych, oddzielnie w obu grupach: eksperymentalnej i kontrolnej.

<sup>11</sup> Bierzemy pod uwagę wynik, gdzie nie założono równości wariancji, ponieważ z Testu Levene'a wynika, iż przy poziomie istotności 0,05 należy odrzucić hipotezę o równości wariancji w obu grupach.

polepszyły się<sup>12</sup>. To ograniczenie dotyczy zresztą wszelkich metod opierających się na badaniu różnic pomiędzy dwiema lub więcej grupami.

Schemat klasyczny nie pozwala również śledzić wpływu wielu zmiennych jednocześnie, lub różnych poziomów bodźca, ani zmienności następstw bodźców w czasie: ich trwałości, słabnięcia i nasilania się. Jako schemat tylko jednej grupy eksperymentalnej i jednej grupy kontrolnej nie pozwala badać uwarunkowań wpływu bodźca, istotnych dla oceny możliwości uogólnień stwierdzonych zależności. Problemy te sprawiły, że powstało wiele przekształceń schematu klasycznego, z których dwa krótko omówiono poniżej.

### **3. Schemat z grupą kontrolną bez pretestu**

Wpływ bodźca w schemacie klasycznym wyrażała różnica między zmianą w grupie eksperymentalnej a zmianą w grupie kontrolnej. Jeśli jednak obie grupy były starannie dobrane i w obu pretest dał jednakowe wyniki (tak było w opisanym wyżej przykładzie), to wpływ bodźca równie dobrze wyraża różnica tylko między wynikami posttestów w obu grupach. Znajomość więc konkretnej wartości zmiennej zależnej przed wprowadzeniem bodźca nie jest konieczna do zmierzenia wpływu bodźca - wystarczy wiedzieć, że wartość ta była jednakowa w obu grupach. Przekonanie to można uzyskać dzięki losowemu doborowi jednostek do poszczególnych grup, albo przez celowe wyrównanie grup ze względu na wartości zmiennych, o których zakładamy, że są skorelowane ze zmienną zależną. Wówczas:  $d' = P_{E2} - P_{K2}$ , jest miarą siły bodźca.

Taki schemat jest bardzo popularny w naukach społecznych. Warto zaznaczyć, że gdy pretest nie jest konieczny, możemy dzięki temu uniknąć wspomnianego wyżej, niechcianego wpływu pretestu na postawy czy zachowania badanych. Jeśli np. informacje o ocenach, w celu wyliczenia średniej, zbierane byłyby u dzieci biorących udział w programie (por. przykład 2) lub dzieci te wiedziałyby o tym, że takie informacje są przez ewaluatorów zbierane, mogłyby pod wpływem tego zmienić swoje zachowanie (np. bardziej się starać i pilniej uczyć), co nie pozostałoby prawdopodobnie bez wpływu na ich wyniki w nauce na koniec roku.

Schemat bez pretestu ma jednak wady, których nie ma schemat klasyczny. Założenie o wstępnym wyrównaniu wartości zmiennej zależnej w grupie eksperymentalnej i kontrolnej ma charakter hipotetyczny. W przypadku, gdy grupy dobiera się losowo i są to grupy liczne,

---

<sup>12</sup> Może wystąpić również efekt odwrotny: dzieci, które "nie dostały się do programu" mogą poczuć się zdemotywowane do dalszej pracy i w efekcie pogorszyć swoje wyniki, czego skutkiem będzie większa różnica pomiędzy wynikami grupy kontrolnej i eksperymentalnej.

nie osłabia to dowodu przyczynowego. Jednak w innych przypadkach trafność tego założenia zależy od teorii, wiążącej wartości mierzonych zmiennych istotnych z nie mierzonymi wartościami zmiennej zależnej. Ponadto, brak pretestu uniemożliwia badanie zmian na poziomie jednostek - umożliwia tylko obserwację zmian na poziomie grup. Nie można więc zbadać procesu, który doprowadził do obserwowanych różnic wewnątrz grupy. Nie docenia się również kierunków i rozmiarów zmiany - jeśli wewnątrz grup zmiany miały kierunki sprzeczne, to znoszą się one i widoczna jest tylko ich wypadkowa. Nie możliwe jest również badanie uwarunkowań zróżnicowania wpływu bodźca, np. szukania odpowiedzi dlaczego jedne dzieci polepszyły swoje wyniki w nauce biorąc udział w programie a inne nie?

#### **4. Schemat wielu bodźców wielowartościowych z kompletną randomizacją**

W opisywanym wcześniej przykładzie badania wpływu programu edukacyjnego na wyniki szkolne dzieci, porównywano sytuację, w której jedynymi alternatywami był udział lub nie w programie (badaliśmy bodziec i brak bodźca). Jednak **ewaluacja wymaga nieraz porównania ze sobą wpływu różnych działań** (różne bodźce) lub różnych wariantów jednego działania (różne poziomy bodźca). Np. jeśli w innym OPS prowadzone byłyby różne, alternatywne formy pomocy dzieciom mającym problemy w nauce, np.: pomoc finansowa dla rodziny na zakup pomocy szkolnych oraz zajęcia indywidualne z korepetytorem w wymiarze 1 i 4 godzin tygodniowo, to interesujące byłoby porównanie wpływu każdej z tych form na wyniki szkolne dzieci objętych daną formą pomocy. W tym celu możemy zastosować schemat kompletnej randomizacji (inaczej schemat grup zrandomizowanych). Polega on na rozlosowaniu jednostek badania na grupy porównawcze, przydzielone poszczególnym rodzajom bodźca, przy czym rodzajów może być więcej niż dwa, a jednym z nich może być również brak oddziaływania. Podstawą pomiaru efektów jest posttest. Metodą zaś analizy wyników jest w tym przypadku analiza wariancji. Analiza wariancji, czyli inaczej zróżnicowania badanych pod względem interesującej nas zmiennej (cechy), polega na podziale całkowitej wariancji wyników eksperymentu na: wariancję między grupami porównawczymi (nazywaną wariancją systematyczną) - czyli tę część zmienności badanej cechy, która wynika z różnic między bodźcami (działaniami), oraz wariancję wewnątrz grup - czyli tę część zmienności badanej cechy, która jest rezultatem oddziaływania wszelkich innych czynników, poza analizowanymi bodźcami (i nazywana jest wariancją błędu). Ograniczeniem stosowania analizy wariancji, podobnie jak testu t-studenta dla porównania średnich, jest jednak to, że grupy porównawcze muszą być dobrane losowo i spełniony powinien być warunek, że rozkłady zmiennej zależnej w grupach są normalne. W poniższej



ramce pokazano przykład zastosowania schematu kompletnej randomizacji i analizy wariancji do zbadania wpływu różnych rodzajów pomocy dzieciom mającym problemy w nauce w pewnym OPS.

**Ramka 3**

**Przykład 3. Zastosowanie schematu kompletnej randomizacji i analizy wariancji do zbadania wpływu różnych rodzajów pomocy dzieciom mającym problemy w nauce w pewnym OPS.**

W pewnym OPS zdecydowano się wprowadzić nowe formy pomocy dzieciom pochodzącym z ubogich rodzin i mającym problemy w nauce. Zaproponowano dwa rodzaje wsparcia: zindywidualizowane korepetycje (z przedmiotów, z którymi dzieci mają największy problem) w dwóch wariantach: 1 lub 4 godziny tygodniowo oraz dofinansowanie zakupu pomocy szkolnych (komputery, zeszyty, podręczniki itp.). Aby zbadać różnice we wpływie każdej z tych form na wyniki szkolne dzieci, zdecydowano się przeprowadzić eksperyment. Rozlosowano 28 dzieci objętych opieką OPS pomiędzy cztery grupy: (1) dzieci, które korzystały z korepetycji 1 godzinę tygodniowo, (2) dzieci korzystające z korepetycji 4 godziny tygodniowo, (3) dzieci, których rodzice otrzymali dofinansowanie na zakup pomocy szkolnych, oraz (4) dzieci, które nie otrzymały żadnego wsparcia<sup>13</sup>. Po 6 miesiącach od wprowadzenia wymienionych form pomocy zdecydowano się na przeprowadzenie posttestu, a więc dokonanie pomiaru średniej ocen dla każdego dziecka z każdej grupy i przeprowadzenie analizy wariancji. Poniższe tabele prezentują wyniki analizy wariancji dokonanej z pomocą programu SPSS<sup>14</sup>.

**Jednoczynnikowa ANOVA**

posttest

	Suma kwadratów	df	Średni kwadrat	F	Istotność
Między grupami	4,507	3	1,502	4,087	,018
Wewnątrz grup	8,821	24	,368		
Ogółem	13,328	27			

<sup>13</sup> Problemy etyczne badań eksperymentalnych zostaną rozważone w dalszej części rozdziału.

<sup>14</sup> Aby przeprowadzić analizę wariancji należy utworzyć dwie zmienne: Zmienną "posttest" i zmienną "grupy" (analogicznie jak w przykładzie 2). Następnie wybrać z menu: Analiza --> Porównanie średnich --> Jednoczynnikowa ANOVA. W okienku ANOVY jako zmienną zależną wybieramy "posttest", zaś jako czynnik zmienną "grupy", w opcji "Post hoc.." wybieramy odpowiedni test (np. T2 Tamhane'a, ponieważ nie musi być w tym przypadku spełniony warunek o jednorodności wariancji w grupach, ).

### Porównania wielokrotne

Zmienna zależna: posttest

Test Tamhane

(I) grupy	(J) grupy	Różnica średnich (I-J)	Błąd standardowy	Istotność	95% przedział ufności	
					Dolna granica	Górna granica
1	2	-,85714	,27355	,054	-1,7249	,0106
	3	-,25000	,34871	,982	-1,3579	,8579
	4	,21429	,32472	,988	-,8090	1,2375
2	1	,85714	,27355	,054	-,0106	1,7249
	3	,60714	,32341	,432	-,4485	1,6628
	4	1,07143*	,29738	,026	,1156	2,0272
3	1	,25000	,34871	,982	-,8579	1,3579
	2	-,60714	,32341	,432	-1,6628	,4485
	4	,46429	,36770	,793	-,6939	1,6225
4	1	-,21429	,32472	,988	-1,2375	,8090
	2	-1,07143*	,29738	,026	-2,0272	-,1156
	3	-,46429	,36770	,793	-1,6225	,6939

\*. Różnica średnich jest istotna na poziomie .05.

Różnica między średnimi ocen dzieci, które otrzymały różne formy pomocy (w tym jej brak), jest istotna statystycznie ( $F(3,24) = 4,087$ ;  $p < 0,018$ ), a więc można stwierdzić, że formy pomocy zastosowane w danym OPS różnie wpływają na wyniki szkolne dzieci. Pozostaje jednak pytanie, które i w jaki sposób? Jak wykazały porównania parami, każdej z form pomocy (porównania wielokrotne), jedynie średnia dzieci korzystających ze zindywidualizowanych korepetycji w wymiarze 4 godzin tygodniowo była istotnie różna (wyższa) od średniej dzieci, które nie korzystały z żadnych form pomocy ( $p < 0,05$ ).

Źródło: opracowanie własne.

Podstawową wadą schematu kompletnej randomizacji jest to, że całe zróżnicowanie między jednostkami składa się na błąd eksperymentalny i obniża istotność wyników. Schemat ten może być więc stosowany przede wszystkim tam, gdzie spodziewane są duże różnice w efektach poszczególnych rodzajów działań (rośnie wówczas wariancja systematyczna) albo tam gdzie mamy do czynienia ze zbiorowościami bardzo jednorodnymi (maleje wtedy wariancja błędu).

## 5. Metody quasi eksperymentalne

W praktyce działań służb społecznych często nie jest możliwe przeprowadzenie eksperymentu. Po pierwsze, celem służb społecznych jest niesienie pomocy wszystkim potrzebującym ludziom, zaś wykonanie eksperymentu wymaga celowego ograniczenia oddziaływania danego rodzaju pomocy jedynie do wybranych osób lub grup (tak jak to było opisane w przykładzie 3). Taka sytuacja stawia eksperymentatora w konflikcie z zasadami etyki zawodowej pracowników służb społecznych. Po drugie, ewaluacja może dotyczyć

działań, które już zostały podjęte (*ewaluacja ex post*) i nie ma możliwości nie tylko wywołania bodźca, ale i dokonania odpowiedniego doboru (w szczególności losowego) jednostek do badania oraz skonstruowania odpowiedniej grupy kontrolnej. Po trzecie, przeprowadzanie eksperymentów stwarza dodatkowe koszty, co w obliczu ograniczonych środków, jakimi dysponują służby społeczne jest istotnym problemem.

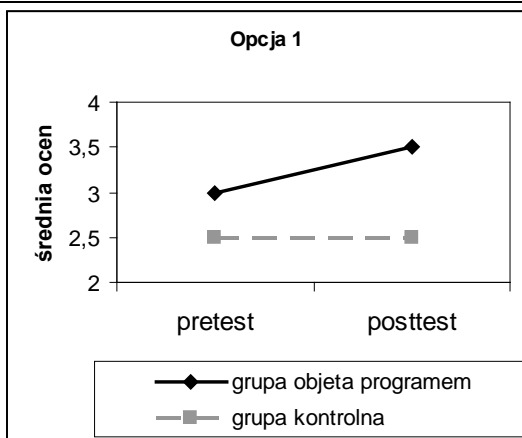
W związku z istnieniem powyższych ograniczeń w realizacji eksperymentów, w praktyce stosuje się również często metody quasi-eksperymentalne. Polegają one przede wszystkim na przybliżaniu warunków badania jak najbliżej warunków eksperymentalnych. Należy jednak wyraźnie podkreślić, że wiarygodność tych badań jest znacznie niższa niż badań eksperymentalnych.

Jedną bardzo popularną metodą quasi-eksperymentalną jest, podobnie jak w eksperymentalnym schemacie klasycznym, porównanie dwóch grup, gdzie jedna to grupa, która została poddana działaniu jakiegoś bodźca a druga to grupa kontrolna - jednak przyporządkowanie do obu grup nie jest losowe lub grupy nie są ujednocicone, tak jak być powinny w schemacie eksperymentalnym. Często grupa, która poddana zostaje działaniu bodźca jest dana, zaś grupę kontrolną dobiera się *ex post*, dążąc oczywiście do tego, by grupa ta była jak najbardziej podobna do grupy eksperymentalnej. Nigdy jednak nie możemy mieć pewności, że grupy te będą porównywalne, zaś każda różnica pomiędzy grupami może mieć wpływ na wynik badania i w skrajnym przypadku może doprowadzić nas do konkluzji, że dane działanie jest skuteczne, gdy faktycznie nie jest lub że nie jest skuteczne, gdy faktycznie jest. W poniższej ramce przedstawiono cztery różne potencjalne rezultaty badania wpływu programu edukacyjnego (omówionego w przykładzie 2) na wyniki szkolne dzieci, w sytuacji gdyby grupy nie były dobrane losowo oraz ich możliwe interpretacje.

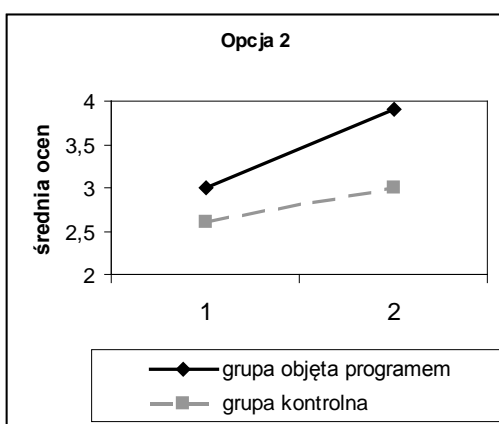
#### **Ramka 4**

#### **Przykład 4. Interpretacje potencjalnych wyników badania wpływu programu edukacyjnego na wyniki szkolne dzieci w sytuacji, gdy podział na porównywane grupy nie jest losowy (grupy nie są odpowiednio ujednocicone).**

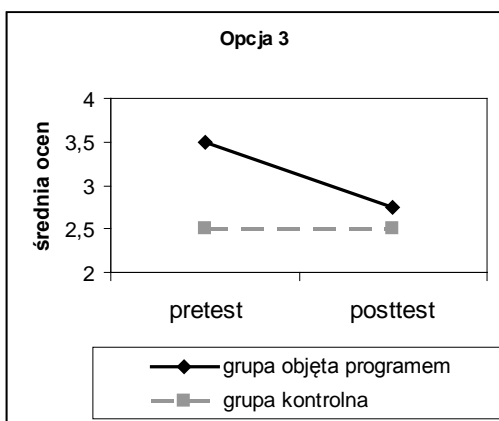
Załóżmy, że w opisanym w przykładzie 2 badaniu wpływu programu edukacyjnego na wyniki szkolne dzieci nie było możliwe dobranie losowe do grup eksperymentalnej i kontrolnej. Zgodnie z założeniami programu, objęto nim dzieci z najuboższych rodzin. Grupę kontrolną skonstruowano spośród pozostałych dzieci, starając się aby ich wyniki szkolne nie odbiegały zbyt od wyników dzieci objętych programem oraz aby posiadały one podobne wartości istotnych cech (jak np. wiek, sytuacja rodzinna, szkoły do których należą dzieci etc.). Nie były to jednak grupy jednolite. Zobaczmy jak można zinterpretować następujące potencjalne rezultaty (porównania średnich ocen z obu grup w preteście i postteście):



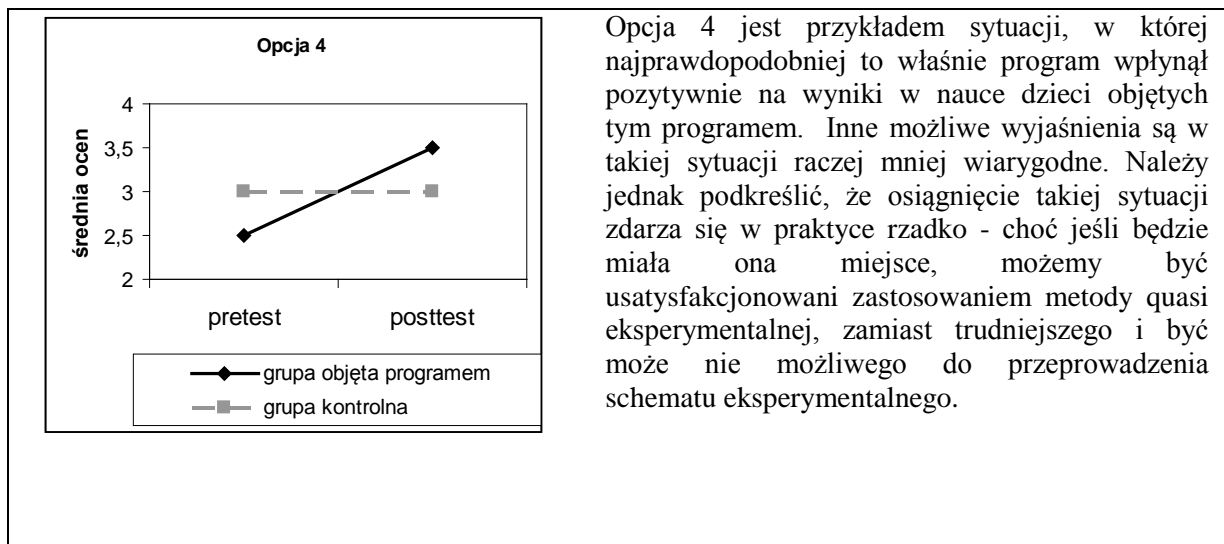
Na pierwszy rzut oka wyniki wydają się wskazywać na to, że program okazał się skuteczny. Jednak ponieważ grupy nie były ujednoczone (losowe) nie można wykluczyć, że grupa objęta działaniem, zareagowała na inny niż program bodziec, na który z kolei nie zareagowała grupa kontrolna. Dla przykładu, jeśli w szkołach, do których uczęszczają dzieci, przyznaje się na koniec roku nagrody tym dzieciom, które osiągnęły z danego przedmiotu minimum ocenę dobrą (4), to dzieci z przeciętnie wyższymi ocenami z grupy poddanej działaniu programu, mogły mieć większą motywację do uczenia się (bliżej było im do zdobycia dobrych ocen) niż dzieci z grupy kontrolnej.



W opcji 2 obserwujemy polepszenie rezultatów w obu grupach, choć poprawa w grupie objętej programem jest większa, co sugeruje, że program był skuteczny. Alternatywnym wyjaśnieniem może być to, że dzieci z przeciętnie wyższymi ocenami prawdopodobnie są zdolniejsze i są w stanie w takim samym czasie dokonać większego postępu w nauce niż dzieci z niższą średnią. W tym przypadku zasadne mogłoby być również alternatywne wyjaśnienie zaprezentowane wyżej dla opcji nr 1. Ponadto grupa, którą objęto działaniem, mogła ulec większej mobilizacji przez sam fakt udziału w programie.



W przypadku opcji 3, wydaje się, że program negatywnie wpłynął na dzieci nim objęte. Tu jednak należałoby wziąć pod uwagę wspomniany na początku rozdziału efekt regresji, czyli sprawdzić, czy przypadkiem szczególnie wysokie oceny w preteście dzieci z grupy objętej programem nie były wynikiem "odchylenia od trendu", zaś niższa średnia w postteście była po prostu powrotem do normy. Podobnie efekt regresji należałoby wziąć pod uwagę, gdyby średnia w preteście dla grupy objętej programem była znacznie niższa niż w grupie kontrolnej, zaś w postteście uległaby znacznej poprawie.



Opcja 4 jest przykładem sytuacji, w której najprawdopodobniej to właśnie program wpłynął pozytywnie na wyniki w nauce dzieci objętych tym programem. Inne możliwe wyjaśnienia są w takiej sytuacji raczej mniej wiarygodne. Należy jednak podkreślić, że osiągnięcie takiej sytuacji zdarza się w praktyce rzadko - choć jeśli będzie miała ona miejsce, możemy być usatysfakcjonowani zastosowaniem metody quasi-eksperymentalnej, zamiast trudniejszego i być może nie możliwego do przeprowadzenia schematu eksperymentalnego.

Na zakończenie należy wyraźnie podkreślić, że stosując metody quasi-eksperymentalne, należy bardzo dokładnie przeanalizować wszystkie alternatywne wyjaśnienia obserwowanych zmian (lub ich braku)<sup>15</sup>.

*Hasła do indeksu*

*Klasyczny eksperyment stochastyczny*

*Pretest*

*Posttest*

*Grupa eksperymentalna*

*Grupa kontrolna*

*Schemat eksperymentalny z grupą kontrolną bez pretestu*

*Schemat eksperymentalny wielu bodźców wielowartościowych z kompletną randomizacją*

*Metody quasi-eksperymentalne*

<sup>15</sup> Na temat metod statystycznej analizy istotności obserwowanych wyników w metodach quasi-eksperymentalnych można przeczytać na stronie: <http://www.socialresearchmethods.net/kb/statnegd.php>