

BADANIA EKONOMICZNE

Paweł Strawiński

Propensity Score Matching

Własności małopróbkowe



Propensity Score Matching



Komitet Redakcyjny serii

Przewodnicząca: prof. dr hab. Urszula Sztandar-Sztanderska

Członkowie: prof. dr hab. Dariusz Dziuba

prof. dr hab. Jacek Kochanowicz

dr hab. Ryszard Kokoszczyński

prof. dr hab. Marek Okólski

prof. dr hab. Krzysztof Opolski

dr hab. Wojciech Otto

prof. dr hab. Jerzy Wilkin

w serii ukazały się m.in.

Michał Brzozowski, Paweł Gierałtowski, Dominika Milczarek, Joanna Siwińska-Gorzela

Institucje a polityka makroekonomiczna i wzrost gospodarczy

Katarzyna Metelska-Szaniawska

*Konstytucyjne czynniki reform gospodarczych w krajach postsocjalistycznych
Studium empiryczne*

Stanisław Kubiela

*Innowacje i luka technologiczna w gospodarce globalnej opartej na wiedzy
Strukturalne i makroekonomiczne uwarunkowania*

Katarzyna Kowalska

Koordynowana opieka zdrowotna. Doświadczenia polskie i międzynarodowe

Wiktor Rutkowski

Współczesne państwo dobrobytu. Ekspansja, kryzys, spory

Łukasz Hardt

Ekonomia kosztów transakcyjnych – geneza i kierunki rozwoju

Ewa Aksman

Redystrybucja dochodów i jej wpływ na dobrobyt społeczny w Polsce w latach 1995–2007

Joanna Tyrowicz

Histeresa bezrobocia w Polsce

Agnieszka Fihel

Płeć a trwanie życia. Analiza demograficzna

Gabriela Grotkowska

Czy Polska importuje bezrobocie? O wpływie handlu zagranicznego na polski rynek pracy w okresie transformacji

Katarzyna Śledziwska

Regionalizm handlowy w XXI wieku. Przesłanki teoretyczne i analiza empiryczna

Tomasz Łyziak

Oczekiwania inflacyjne konsumentów: pomiar i testowanie

Paweł Strawiński

Propensity Score Matching

Własności małopróbkowe



UNIwersytet Warszawski
**Wydział Nauk
Ekonomicznych**

Warszawa 2014

Recenzent

dr hab. Adam Szulc

Projekt okładki i stron tytułowych

Frycz/Wicha/corporate design

Edwin Radzikowski

Redaktor prowadzący

Małgorzata Yamazaki

Redakcja

Krystyna Dziewanowska-Stefańczyk

Skład i łamanie

Fixpoint

© Copyright by Wydawnictwa Uniwersytetu Warszawskiego, Warszawa 2014

ISBN 978-83-235-1696-5 (PDF)

Wydawnictwa Uniwersytetu Warszawskiego

00-497 Warszawa, ul. Nowy Świat 4

www.wuw.pl; e-mail: wuw@uw.edu.pl

Dział Handlowy WUW: tel. (48 22) 55-31-333; e-mail: dz.handlowy@uw.edu.pl

Księgarnia internetowa: www.wuw.pl/ksiegarnia

Wydanie I

Spis treści

Wprowadzenie	7
1. Efekty oddziaływania i łączenie danych	13
1.1. Wprowadzenie	13
1.2. Fundamenty analizy kontrfaktycznej	14
1.3. Łączenie danych	21
1.3.1. Statystyczna metoda łączenia	24
1.3.2. <i>Propensity Score Matching</i>	29
1.4. Algorytmy łączenia używające <i>propensity score</i>	43
1.4.1. Dobieranie najbliższego sąsiada	45
1.4.2. Ważenie odwrotnością prawdopodobieństwa	47
1.4.3. Kilka słów o weryfikacji założeń	48
1.5. Podsumowanie	52
2. Własności małopróbkowe estymatorów metody łączenia	56
2.1. Wprowadzenie	56
2.1.1. Wybór prac o własnościach małopróbkowych PSM	57
2.1.2. Wymuszanie wspólnej części przedziału określoności	60
2.1.3. Opis symulacji	62
2.1.4. Rozmiar i moc symulacji	69
2.2. Szacowanie przeciętnych efektów oddziaływania	75
2.2.1. Łączenie 1 do 1	76
2.2.2. Ważenie odwrotnością prawdopodobieństwa	90
2.3. Szacowanie przeciętnych efektów oddziaływania wobec jednostek poddanych oddziaływanii	95
2.3.1. Łączenie 1 do 1	96
2.3.2. Ważenie odwrotnością prawdopodobieństwa	107
2.4. Podsumowanie wyników symulacji	111

3. Zastosowania <i>Propensity Score Matching</i>	115
3.1. Wprowadzenie	115
3.2. Oszacowanie wpływu szkoleń zawodowych na zarobki	116
3.3. Opłacalność strategii dochodowych na polskiej wsi	126
3.4. Ocena skuteczności subsydiowanych miejsc pracy	137
3.5. Prywatne a publiczne szkoły w Hiszpanii	145
3.6. Podsumowanie	151
Summary	155
Aneks. Angielsko-polski słownik terminów specjalistycznych	159
Bibliografia	163
Spis rysunków	169

Wprowadzenie

W ostatnich latach metody mikroekonometryczne zdobywają coraz większą popularność i uznanie badaczy. Jest to związane przede wszystkim z rozwojem technik gromadzenia i przetwarzania danych. Mikroekonometria to grupa metod przeznaczona do analizy mikrodanych. Współcześnie mikrodane, czyli dane jednostkowe, są gromadzone w wielu dziedzinach gospodarki takich jak finanse, ubezpieczenia, rynek pracy czy badanie zachowania gospodarstw domowych.

Jednym z obszarów analiz ilościowych prowadzonych na poziomie mikro jest analiza skuteczności działań podejmowanych przez instytucje publiczne. Dotyczy to przede wszystkim różnych form wsparcia takich jak szkolenia podnoszące kwalifikacje pracowników, szkolenia dla osób bezrobotnych czy programy wsparcia dla firm. Z reguły takie działania są skierowane do specyficznych grup odbiorców. Występujące zjawisko selekcji lub samoselekcji próby uniemożliwia przeprowadzenie ilościowej oceny skuteczności tych działań z wykorzystaniem klasycznych narzędzi ekonometrycznych (Strawiński 2007).

Metoda łączenia danych jest narzędziem pozwalającym na ocenę skuteczności podejmowanych działań o charakterze programów. Jest to grupa metod, wywodząca się z nauk eksperymentalnych i przeniesiona na grunt dziedzin nauki, w których prowadzenie w pełni kontrolowanych eksperymentów nie jest możliwe albo jest w znacznym stopniu ograniczone. Na przykład, nie jest możliwe przeprowadzenie eksperymentalnego badania skuteczności leku, który może zagrażać zdrowiu lub życiu (Szulc 2012).

Łączenie danych pozwala uzyskać odpowiedź na pytanie, „co by się stało, gdyby?”. Nie jest ono pytaniem nowym ani specyficznym

dla nauk ekonomicznych. W naukach eksperymentalnych badacze poszukują na nie odpowiedzi, dokonując serii eksperymentów. W dziedzinach, w których nie jest to możliwe, wykorzystuje się dostępny aparat matematyczny do nadania właściwości zbioru eksperymentalnego danym niepochodzącym z eksperymentu. Jedną z takich metod jest łączenie danych. Umożliwia ona przeprowadzenie analizy kontrfaktycznej. Jednym ze sposobów łączenia danych, który w ostatnich latach uzyskał uznanie badaczy, jest łączenie według prawdopodobieństwa (ang. *Propensity Score Matching*). Zastosowanie metody nie ogranicza się do nauk ekonomicznych, jest ona z powodzeniem wykorzystywana w medycynie, naukach politycznych czy naukach rolniczych.

Niniejsza książka jest pogłębionym studium własności statystycznych estymatorów metody *Propensity Score Matching*, stosowanych do szacowania przeciętnych efektów oddziaływania. Asymptotyczne właściwości estymatorów zostały już przez badaczy opisane, brakuje natomiast badań dotyczących prób o niewielkiej liczbie obserwacji. Niniejsza książka jest próbą wypełnienia tej luki. Efekty oddziaływania można szacować, wykorzystując wiele sposobów. Dwoma najczęściej wykorzystywanymi miarami są przeciętny efekt oddziaływania oraz przeciętny efekt oddziaływania wobec jednostek poddanych oddziaływaniu. Pierwsza wskazuje, jaki jest efekt działania programu wobec przeciętnej jednostki w populacji. Druga miara opisuje efekt działania programu wobec jednostek w tym programie uczestniczących. W kontekście szacowania tych wielkości zostaną ukazane właściwości statystyczne estymatorów uzyskiwanych z wykorzystaniem dwóch metod bazujących na oszacowaniach *propensity score*, czyli prawdopodobieństwa poddania oddziaływaniu. Wspomnianymi metodami będą łączenie danych według prawdopodobieństwa oraz ważenie odwrotnością oszacowania prawdopodobieństwa.

Pewną trudnością, która stała przed autorem, jest brak ugruntowanej terminologii w języku polskim dotyczącej metod łączenia danych. Jedyną znaną autorowi pozycją o charakterze akademickim, w której tematyka łączenia danych jest omawiana, jest podręcznik *Mikroekonometria* pod redakcją prof. dr. hab. Marka Gruszczyńskiego. Kilka lat temu, przed powstaniem pierwszego wydania podręcznika, próbo-

waliśmy uzgodnić jednolite tłumaczenia anglojęzycznych terminów. W wielu przypadkach się to udało, jednak drobne różnice pozostały. Przykładem takiej rozbieżności jest tłumaczenie na język polski pojęcia *matching estimation*. W podręczniku *Mikroekonometria* zaproponowano „estymacja przez dopasowanie” (rozdział 9 autorstwa Adama Szulca), ja proponuję „estymacja przez łączenie”. Obie propozycje są dosłownymi tłumaczeniem pojęcia *matching*. Oznacza ono zarówno dopasowywanie, dobieranie, jak i łączenie w pary. Wydaje mi się, że termin „łączenie” lepiej oddaje charakter pojęcia, gdyż jego istotą jest łączenie podobnych obserwacji, najczęściej w pary. Ponadto w ekonometrii słowo „dopasowanie” jest kojarzone ze stopniem dopasowania modelu do danych empirycznych. Aby uniknąć niejasności terminologicznej, na końcu niniejszej książki jest umieszczony załącznik, w którym zestawiono stosowane tłumaczenia na język polski terminów anglojęzycznych. Dodatkowo w celu uniknięcia zamieszania, w książce obok określeń polskich pojawiają się ich anglojęzyczne odpowiedniki.

Osobnego komentarza wymaga wyjaśnienie tytułu książki. Jego pierwszy człon jest w języku angielskim, nie ma bowiem utrwalonego w literaturze tłumaczenia go na język polski. Moje tłumaczenie „łączenie według prawdopodobieństwa” stara się oddać istotę metody, jednak nawet dla specjalistów w dziedzinie mikroekonometrii mogłoby być niezrozumiałe. W pracy oba pojęcia będą pojawiać się wymiennie, chociaż *Propensity Score Matching* częściej. Ta sama uwaga dotyczy pojęcia *propensity score* i jego tłumaczenia – „prawdopodobieństwo oddziaływania”.

Na książkę składają się trzy rozdziały. Zadaniem pierwszego jest wprowadzenie czytelnika w świat efektów oddziaływania i łączenia danych. Ma on charakter teoretycznego przeglądu najważniejszych prac z literatury dotyczącej analizy kontrfaktycznej i łączenia danych. Zaprezentowano fundamenty analizy kontrfaktycznej, które wywodzą się z prac Neymana (1923a i 1923b) i Fishera (1925) na temat randomizacji. Opisano dwa podstawowe efekty oddziaływania oraz warunki ich identyfikacji. Opierając się na fundamentalnej pracy Rosenbauma i Rubina (1983), omówiono metodę łączenia obserwacji oraz propozycję autorów wykorzystania *propensity score* do łączenia obserwa-

cji. W dalszej części rozdziału przedstawiono założenia i statystyczne właściwości metody *Propensity Score Matching* w dużych próbach. Fragment tej części bazuje na mojej rozprawie doktorskiej. Wartością dodaną jest zebranie w jednym miejscu wyników najważniejszych prac z ostatnich 15 lat. Rozdział zamyka prezentacja różnych algorytmów używanych do łączenia danych oraz narzędzi diagnostycznych pozwalających ocenić jakość uzyskanego łączenia. Omówiono różne warianty algorytmu dobierania najbliższego sąsiada – jego zaletą jest szybkość, a wadą niejednoznaczność. Innym algorytmem jest ważenie odwrotnością oszacowania prawdopodobieństwa, którego jakość zależy m.in. od jakości przyjętego modelu dla prawdopodobieństwa. Rozdział kończy prezentacja dyskusji naukowej, która toczyła się na początku XXI wieku i dotyczyła zagadnienia, czy wykorzystując metody łączenia według prawdopodobieństwa, można odtworzyć wyniki eksperymentalne.

W rozdziale drugim zaprezentowano wyniki badań symulacyjnych nad właściwościami estymatorów metody łączenia według prawdopodobieństwa w małych próbach. Własności metod statystycznych w małych próbach są kluczowe w praktycznym zastosowaniu narzędzi ekonometrycznych. W konkretnych problemach badawczych, np. w pomiarze efektu interwencji na rynku pracy, badacz zazwyczaj staje przed oceną ilościowego efektu oddziaływania na podstawie badania o charakterze pilotażowym. Cechą charakterystyczną badań pilotażowych jest ich niewielka skala, a zatem niewielka liczebność próby. Ważna jest więc informacja, czy właściwości statystyczne metod łączenia, które udowodniono, korzystając z twierdzeń granicznych, są również zachowane w próbach o liczebnościach zbliżonych do liczebności prób rzeczywistych. W literaturze brakuje ogólnych prac dotyczących własności estymatorów metody *Propensity Score Matching* w małych próbach. W rozdziale zaprezentowano wyniki własnych oryginalnych badań symulacyjnych. Struktura eksperymentalnego planu symulacyjnego jest analogiczna do moich wcześniejszych prac. Wykorzystany plan eksperymentalny oraz część prezentowanych wyników dla estymatorów przeciętnych efektów oddziaływania wobec jednostek poddanych oddziaływaniu zostały wcześniej opublikowane przeze mnie w artykułach. Cechą odróżniającą uzyskane wyniki od wcześniejszych rezultatów jest

cel badania oraz jego dużo szerszy zakres. Analizowano dwa sposoby szacowania efektów oddziaływania oraz wpływ różnych technik wymuszających spełnienie założeń metody *Propensity Score Matching* na statystyczne własności estymatorów dla efektów oddziaływania. Analizowano standardowe miary jakości: obciążenie, wariancję i pierwiastek błędu średniokwadratowego estymatora. Nowym elementem niewystępującym we wcześniejszych pracach są analizy rozmiaru procedury łączenia i mocy testów istotności estymatora oszacowanej wartości efektu oddziaływania.

Celem trzeciego rozdziału jest próba prezentacji zastosowań i właściwości statystycznych omawianych metod zastosowanych do analizy rzeczywistych zbiorów danych. W rozdziale zaprezentowano cztery przykłady. Pierwszy z nich jest oryginalnym badaniem pokazującym, w jaki sposób można wykorzystać dostępne dane pochodzące z Badania Aktywności Ekonomicznej Ludności (BAEL) do oceny skuteczności dodatkowych szkoleń podnoszących kwalifikacje zawodowe oraz ich wpływu na sytuację uczestniczących w nich pracowników. Pozostałe przykłady zostały zaadaptowane z literatury. Drugi z nich powstał na podstawie artykułu, którego jestem współautorem. W przykładzie są porównane relatywne korzyści dla gospodarstw wiejskich z wyboru strategii dywersyfikacji albo specjalizacji działalności. Trzeci przykład pochodzi z mojego własnego artykułu. Są w nim analizowane dane dotyczące amerykańskiego programu subsydiowanych miejsc pracy. Celem badania jest oszacowanie efektu brutto programu, czyli efektu dla beneficjenta bez uwzględniania kosztu programu. Ostatni przykład został zaadaptowany z artykułu Vanderberghe’a i Robina (2004). Celem artykułu jest porównanie wpływu umieszczania dzieci w szkołach prywatnych na ich osiągnięcia edukacyjne.

W każdym z przykładów podjęto próbę oszacowania przeciętnego efektu oddziaływania (ATE) oraz przeciętnego efektu oddziaływania wobec jednostek poddanych oddziaływaniu (ATT) metodą łączenia według oszacowanych wartości *propensity score* oraz metodą przeważania według odwrotności oszacowań *propensity score*. Wskazano na dwa ważne wyniki. Po pierwsze, wykorzystanie metod łączenia danych przyczynia się do zbilansowania rozkładów cech w grupie

eksperymentalnej i dobranej do niej grupie kontrolnej. Po drugie, nie istnieje jedna uniwersalna i zawsze skuteczna metoda wykorzystania wektorów prawdopodobieństwa oddziaływania w małych próbach.

Rozdział 1

Efekty oddziaływania i łączenie danych

1.1. Wprowadzenie

Metoda łączenia według prawdopodobieństwa jest jednym ze sposobów na szacowanie wielkości efektów oddziaływania. Trudno jest zatem opisać właściwości tej metody estymacji bez odwoływania się do kontekstu, w jakim znajduje ona zastosowanie. W tym celu zostaną zaprezentowane fundamenty analizy kontrfaktycznej. Stanowi ona ważne narzędzie wspomagające ilościową ewaluację programów. Następnie zostaną omówione metody łączenia danych oraz metodologiczne i statystyczne uwarunkowania wykorzystania łączenia według prawdopodobieństwa. Rozdział zamyka dyskusja, czy wykorzystując metody łączenia według prawdopodobieństwa, można odtworzyć kontrolowane randomizowane eksperymenty.

Za idealną metodę badawczą, wyznaczającą złoty standard nauki, są uznawane w pełni kontrolowane randomizowane eksperymenty. Są one domeną nauk eksperymentalnych takich jak fizyka, biologia czy medycyna. W warunkach laboratoryjnych obserwowane jest zachowanie jednostek przydzielonych do grupy eksperymentalnej i identycznych z nimi jednostek przydzielonych do grupy kontrolnej. Na podstawie porównania reakcji na oddziaływanie zaaplikowane jednostkom w grupie eksperymentalnej z zachowaniem jednostek w grupie kontrolnej są wyciągane wnioski dotyczące efektów oddziaływania. W naukach społecznych przeprowadzenie kontrolowanych eksperymentów na szeroką skalę nie jest możliwe. Najbardziej wiarogodne i uznane wyniki w tej dziedzinie pochodzą z badań, w których stosowano randomizację (LaLonde 1986).

Termin „randomizacja” wprowadzili do statystyki niezależnie Neyman (1935) i Fisher (1935). Fisher opisuje randomizację jako „uza-

sadnioną podstawę” wnioskowania i „podstawy dla poprawności testu [statystycznego]”. Randomizacją nazywamy losowe przydzielenie jednostek do grupy eksperymentalnej i kontrolnej. Zatem nie każdy kontrolowany eksperyment jest randomizowany. Nadzór nad warunkami przeprowadzania eksperymentu i manipulowanie nimi nie musi zapewniać efektu randomizacji. Na przykład podczas prób klinicznych nowatorskie leki są serwowane pacjentom cierpiącym na typ choroby, na który lek ma oddziaływać. Zdanie odwrotne, że każdy randomizowany eksperyment jest kontrolowany, również nie jest prawdziwe. Dobrym przykładem jest znany amerykański program wsparcia osób bezrobotnych *National Supported Work Demonstration*, w którym uczestników w sposób losowy przydzielano do różnych form aktywizacji zawodowej. Ponieważ program był interwencją na rynku pracy, nie wszystkie jego efekty były kontrolowane, np. nie można było oddzielić jednostek poddanych różnym formom aktywizacji w sposób zapewniający brak kontaktu między nimi.

W celu przeprowadzenia skutecznej randomizacji badanie musi zostać odpowiednio zaprojektowane. Jednostki poddane eksperymentowi nie muszą być homogeniczne, próba eksperymentalna i grupa kontrolna nie muszą być próbami losowymi z populacji, ale jednostki muszą być w sposób losowy przydzielone do grupy poddanej oddziaływaniu i grupy kontrolnej (Rosenbaum 2002). Jednak w niektórych przypadkach badacze poszukują uzasadnienia empirycznego dla zjawisk, które nie były głównym przedmiotem zainteresowania badania, wykorzystując zgromadzone dane. W takiej sytuacji metody łączenia pozwalają na przekształcenie danych pochodzących z badań ankietowych w zbiór danych spełniający wszystkie warunki, jakim podlegają dane eksperymentalne. W literaturze w ten sposób przekształcone zbiory określa się mianem danych quasi-eksperymentalnych.

1.2. Fundamenty analizy kontrfaktycznej

Randomizacja jest uważana za metodę wyznaczającą standard oceny wpływu efektu oddziaływania danego czynnika czy wprowadzonej interwencji na rezultat. Losowe przydzielenie oddziaływania oznacza,

że przydział jednostki do grupy eksperymentalnej albo kontrolnej odbywa się w sposób niezależny od wyniku oddziaływania (Heckman i Vytlačil 2007). Zapewnia on, że nie będą występowały inne czynniki jednocześnie wpływające na przydzielenie oddziaływania i wynik jego działania (ang. *confounders*). Dzieje się tak, gdyż wynik działania jest niezależny od przydziału jednostki do konkretnej grupy. Zatem efekt oddziaływania może zostać oszacowany poprzez proste porównanie jednostek poddanych oddziaływaniu i jemu nie poddanych (Neyman 1923a i 1923b; Pearl 2000).

W naukach społecznych, w odróżnieniu od nauk eksperymentalnych, przeprowadzenie kontrolowanego eksperymentu i wykorzystanie randomizacji w większości przypadków nie jest możliwe. W pewnych sytuacjach nawet w naukach eksperymentalnych przeprowadzenie eksperymentu może nie być możliwe ze względu na ograniczenia prawne lub powody etyczne. Dobrym przykładem są ograniczenia w testowaniu klinicznym leków o znacznych efektach ubocznych. Z drugiej strony ocena efektu działania czynnika, np. programu czy prowadzonej polityki wymaga informacji, w jaki sposób jednostka zachowałaby się, gdyby nie została poddana oddziaływaniu tego czynnika.

Standardowe ramy analizy kontrfaktycznej i model potencjalnych wyników są przez różnych autorów różnie przypisywane: Neymanowi (1923a i 1923b), Fisherowi (1925), Royowi (1951), Quandtowi (1972) lub Rubinowi (1973a). Neyman (1923a i 1923b) w pracy dotyczącej randomizowanych eksperymentów sformułował problem w języku statystycznym. Następnie Roy (1951) zauważył, że w badaniach ankietowych (obserwacyjnych) obserwowane są wyłącznie najlepsze dla rozpatrywanych jednostek wybory. Uzależnił on podział jednostek na poddane i niepoddane oddziaływaniu od potencjalnych wyników działania. Matematyczne powiązanie obserwowanego wyniku oddziaływania ze średnią ważoną wyniku dla grupy eksperymentalnej i kontrolnej, gdzie wagami są udziały jednostek z grupy eksperymentalnej i kontrolnej w próbie, jest efektem pracy Quandta (1972). Z kolei Rubin (1973a) stworzył regułę podziału jednostek, która jest niezależna od potencjalnego wyniku. Innymi słowy, przydzielenie do jednostek poddanych albo niepoddanych oddziaływaniu nie zależy od potencjalnego wyniku działania bądź braku

działania. W celu oszacowania efektu oddziaływania zaproponował on łączenia podobnych obserwacji i porównanie ich wyników. Ten sposób analizy jest często w literaturze anglojęzycznej przypisywany Royowi (1951) oraz Rubinowi (1973b) i opisywany jako model Roya–Rubina.

Randomizacja i oddziaływanie ściśle wiążą się z pojęciem przyczynowości (Strawiński 2007). Relacja przyczynowo-skutkowa we współczesnej statystyce i ekonometrii jest przedstawiana z zastosowaniem notacji wykorzystującej wyniki kontrfaktyczne (ang. *counterfactual*). W przypadku badań nieeksperymentalnych są one wartościami hipotetycznymi, a zatem nie można tych wyników obserwować w rzeczywistości. Podstawowymi elementami analizy są jednostki, oddziaływanie i potencjalne efekty oddziaływania, czyli potencjalne wyniki. Najprostszym typem oddziaływania, jakiemu mogą być poddane jednostki, jest tzw. oddziaływanie binarne, w którym czynnik oddziałuje na jednostkę albo na nią nie oddziałuje.

Notacja odnosząca się do potencjalnych wyników została zaproponowana w pracach Neymana (1923a i 1923b) i Fishera (1925), dotyczących randomizowanych eksperymentów. Niech zmienna losowa P opisuje dwa wyróżnione stany: poddania i niepoddania oddziaływaniu. Stan poddania oddziaływaniu oznaczamy $P_i = 1$ i mówimy wówczas, że jednostka należy do grupy eksperymentalnej (ang. *treated group*). Stan niepoddania oddziaływaniu $P_i = 0$ nazywamy przynależnością jednostki do grupy niepoddanej oddziaływaniu (ang. *non-treated*) lub grupy kontrolnej (ang. *control treatment*). Oczywiście role stanu $P_i = 0$ i $P_i = 1$ można odwrócić (Heckman, LaLonde i Smith 1999). Zatem wybór stanu $P_i = 1$ często jest arbitralny. W przypadku analiz prowadzonych w naukach społecznych stan $P_i = 0$ jest w naturalny sposób przypisywany brakowi interwencji. Z tego powodu w literaturze ekonomicznej często zamiast o pojęciu grupy eksperymentalnej mówi się o grupie jednostek poddanych działaniu programu bądź polityki. Dodatkowo przyjmijmy następujące oznaczenia. Niech Y_{1i} będzie wynikiem działania w sytuacji, gdy jednostka i jest poddana oddziaływaniu lub należy do grupy eksperymentalnej, a Y_{0i} to wynik, gdy pozostaje poza wpływem działania lub należy do grupy pozostającej poza programem. W tym miejscu warto dokonać rozgraniczenia dwóch pojęć: grupa