

METODOLOGICZNE INSPIRACJE 2016

**BADANIA ILOŚCIOWE W NAUKACH SPOŁECZNYCH
WYZWANIA I PROBLEMY**

SZCZEGÓŁOWY PROGRAM KONFERENCJI

Jabłonna, wrzesień 2016

Konferencja *Metodologiczne Inspiracje 2014: Badania Ilościowe w Naukach Społecznych — Problemy i Wyzwania* jest organizowana przez Instytut Filozofii i Socjologii PAN oraz Komitet Socjologii PAN.

Środa, 28. września 2016 r., godz. 15.45 – 18.15

SESJA 1

Przewodniczący: Zbigniew Karpiński

Instytut Filozofii i Socjologii PAN

Coś z niczego czyli czego możemy się dowiedzieć o sieciach społecznych na podstawie danych sondażowych

Michał Bojanowski

Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, Uniwersytet Warszawski

Badania sieci społecznych opierają się zazwyczaj na (1) danych zbieranych w sposób wyczerpujący w określonej zbiorowości aktorów lub (2) na danych zbieranych w sposób sondażowy za pomocą narzędzi takich jak na przykład tzw. „generator imion”. Ideałem są dane zebrane w sposób (1), ale jest to przeważnie bardzo kosztowne, a nierzadko niemożliwe. Z kolei sposób (2) dostarcza tylko informacji o respondentach oraz osobach, z którymi respondenci pozostają w interesującej nas relacji. Jednocześnie, udzielenie odpowiedzi na ważne pytania badawcze (dotyczące np. rozprzestrzeniania się informacji, zachowań, chorób, opinii itp.) wymagają znajomości kompletnej sieci społecznej (ad 1.), a nie tylko zbioru sieci egocentrycznych (ad 2.). Prezentujemy i dyskutujemy nowe podejścia oparte na modelach wykładniczych grafów losowych (ERGM), które pozwalają na podstawie danych egocentrycznych rekonstruować kompletne sieci społeczne w interesującej nas zbiorowości.

Ogólnopolskie badania ekonomicznych losów absolwentów uczelni – pierwszy krok ku powszechnemu wykorzystaniu zasobów administracyjnych państwa do celów naukowych

Mikołaj Jasiński, Tomasz Zając, Marek Bożykowski, Marek Styczeń

Pracownia Ewaluacji Jakości Kształcenia, Uniwersytet Warszawski

Dane administracyjne — przez administrację w ramach wykonywania jej statutowych obowiązków — są wciąż dość rzadko wykorzystywane w naukach społecznych w Polsce.

Doświadczenia badawcze zespołu Pracowni Ewaluacji Jakości Kształcenia Uniwersytetu Warszawskiego potwierdziły wnioski badaczy innych krajów wskazujące na wiele zalet wykorzystania danych z rejestrów administracyjnych do badań społecznych. Wśród najważniejszych należy wymienić dostęp do danych na temat całych populacji, rzetelność gromadzonych informacji oraz relatywnie niskie koszty prowadzenia badań wynikające z braku konieczności zbierania danych. Kluczową kwestią przy udostępnianiu informacji z systemów administracyjnych jest stworzenie możliwości łączenia informacji z różnych rejestrów. Rozwiązanie takie radykalnie zwiększa i tak znaczną wartość danych administracyjnych jako źródła do analiz statystycznych — zarówno dla badaczy, jak i twórców polityk publicznych.

Badania oparte na danych gromadzonych przez administrację mają też swoje ograniczenia związane m.in. z regulacjami dotyczącymi ochrony prywatności, czy ograniczonym wpływem badaczy na sposób zbierania danych przez administrację i zakres gromadzonych informacji.

Przykładem badania wykorzystującego potencjał danych administracyjnych jest *Ogólnopolski monitoring ekonomicznych losów absolwentów*. Badanie przeprowadziliśmy na zlecenie Ministerstwa Nauki i Szkolnictwa Wyższego. Celem projektu była analiza sytuacji zawodowej absolwentów szkół wyższych w Polsce na podstawie danych pochodzących z rejestrów Zakładu Ubezpieczeń Społecznych oraz systemu POL-on. Obok dostarczenia twórcom polityki edukacyjnej, uczelniom i kandydatom informacji na temat sytuacji ekonomicznej absolwentów uczelni, celem naszych działań jest stworzenie zespołom badawczym możliwości prowadzenia analiz bazujących na danych administracyjnych. Obecnie zespół PEJK prowadzi prace nad rozszerzeniem zakresu takich analiz — zarówno ze względu na inne obszary badań, jak i kolejne rejestry administracyjne.

Referat składać się będzie z czterech części. Po pierwsze, przedstawimy genezę projektu, w tym rozwiązania prawne oraz organizacyjne, które umożliwiły jego realizację. Po drugie, omówimy podstawy metodologiczne badania, w tym źródła danych, konstrukcję wykorzystywanych wskaźników oraz ich własności i interpretację. W szczególności przedstawimy koncepcję tzw. wskaźników względnych, które przy analizie skuteczności poszukiwania pracy oraz ocenie uzyskiwanych wynagrodzeń biorą pod uwagę sytuację na lokalnych rynkach pracy. Po trzecie, zaprezentujemy wyniki zakończonego niedawno pierwszego pomiaru ekonomicznych losów absolwentów. Omówimy analizy czasu poszukiwania pracy, stabilności zatrudnienia oraz uzyskiwanych wynagrodzeń. Po czwarte, zasygnalizujemy projekty przygotowywane obecnie przez nasz zespół, które wykraczają poza zagadnienia związane z badaniem losów absolwentów, a nawet poza socjologię.

Badania treści w cyberprzestrzeni. Big Data jako wyzwanie dla socjologii

Paweł Matuszewski

I Wydział Nauk Historycznych i Społecznych Uniwersytet Kardynała Stefana Wyszyńskiego

Pod wpływem rozwoju technologicznego coraz więcej aktywności społecznych przenoszonych jest do świata wirtualnego. Referat będzie dotyczył badań nad przekonaniem zbiorowymi formułowanymi w cyberprzestrzeni. Podstawowym problemem w tego typu analizach jest ilość produkowanych treści, którą nieraz liczy się w gigabajtach danych. Sprawia to, że praktycznie niemożliwym jest, aby jeden badacz zapoznał się z całym dostępnym materiałem jakościowym. Pojawia się zatem konieczność zastosowania innych strategii badawczych, niż te wykorzystywane m.in. w analizie wywiadów pogłębionych. Rozwiązaniem jest zastosowanie statystycznych metod analizy tekstu.

W referacie zaprezentowany zostanie proces analizy dużego zbioru danych tekstowych od momentu pozyskania materiału badawczego przez analizę statystyczną po sformułowanie wniosków socjologicznych. Materiałem badawczym będzie kilkadziesiąt tysięcy komentarzy internautów zebranych na oficjalnych facebookowych profilach wszystkich ogólnopolskich komitetów wyborczych w wyborach parlamentarnych w Polsce w 2015 r.

Międzynarodowe badania sondażowe na temat korupcji w Europie: reprezentacja krajów i ex-post harmonizacja

Ilona Wysmułek

Szkoła Nauk Społecznych, Instytut Filozofii i Socjologii PAN

Celem mojego referatu jest przedstawienie zalet i ograniczeń związanych z posługiwaniem się międzynarodowymi sondażami przy badaniu korupcji. W swojej analizie skupiam się na badaniach przeprowadzonych w Europie w latach 1989-2013, zawierających pytania na temat zachowań korupcyjnych oraz percepcji korupcji. Zebrany materiał na temat korupcji pochodzi m.in. z takich międzynarodowych badań, jak *Eurobarometer*, *Life in Transition Survey*, *Global Corruption Barometer*, *Crime Victims Survey*, *European Social Survey*, *International Social Survey Program*, czy *World Values Survey*. W sumie obejmuje on 63 fale z 19 projektów badawczych. W swoim badaniu główną uwagę zwracam na zakres tematyczny międzynarodowych kwestionariuszy na przykładzie stawianych pytań o korupcję. Wyniki badań wskazują na wzrost zainteresowania problemem korupcji w międzynarodowych badaniach sondażowych. Jednocześnie to zainteresowanie jest nierównomiernie duże w stosunku do poszczególnych krajów. W wystąpieniu planuję poruszyć zagadnienie nierówności reprezentacji krajów w badaniach nad korupcją oraz ich możliwe przyczyny, a także postawić pytanie na temat możliwości ex-post harmonizacji pytań sondażowych.

Model pracowniczych przewag informacyjnych w organizacji

Patrycja Antosz

Instytut Socjologii, Uniwersytet Jagielloński

Umowa o pracę, zgodnie z którą, niezależnie od obciążenia zadaniowego, pracownicy zobowiązani są przepracować określoną ilość czasu, pozostaje jedną z najbardziej popularnych form zatrudnienia. Faktycznie pracownicy poświęcają średnio około dwóch godzin z ośmiogodzinnego dnia pracy na czynności niezwiązane z zatrudnieniem (Paulsen, 2015). Zgodnie z teorią agencji (*principal-agent problem*), pracownicy mogą sobie na to pozwolić, gdyż posiadają dwa rodzaje przewag informacyjnych nad swoimi przełożonymi. Z jednej strony posiadają pełną informację dotyczącą swoich kompetencji, z drugiej strony dotyczącą faktycznej ilości pracy, jaką wykonują. W literaturze ekonomicznej problemy te znane są pod pojęciami negatywnej selekcji (*adverse selection*) (Akerlof, 1970) oraz pokusy nadużycia (*moral hazard*) (Holmstrom, 1979). W psychologii organizacji podobne zjawiska opisywane są pod hasłami wykonania maksymalnego oraz wykonania typowego (Sackett i in., 1988). Wykonanie maksymalne występuje w krótkotrwałej sytuacji ewaluacyjnej, w której pracownik jest świadomy procesu oceny oraz posiada jasne instrukcje dotyczące wykonania zadania. Dzięki wysokiej motywacji pracownika wykonanie maksymalne traktowane jest jako wskaźnik kompetencji. Natomiast wykonanie typowe postrzegane jest raczej w kategoriach osobowości pracownika, w szczególności jego sumienności. W socjologii organizacji wykonanie typowe ujmowane jest w większym stopniu jako efekt kultury organizacyjnej wytworzonej na drodze interakcji przełożonych oraz pracowników, a nie różnic indywidualnych. Konformizm w działaniach jest produktem komunikacji. Podwładni podejmują decyzje na podstawie tego, co robią współpracownicy, gdyż zachowanie innych

jest najpewniejszym źródłem informacji pozwalającej na maksymalizację użyteczności.

Na podstawie teorii ekonomicznych, psychologicznych oraz socjologicznych, a także własnych badań jakościowych w postaci wywiadów pogłębionych z pracownikami oraz menadżerami stworzony został model agentowy pracowniczych przewag informacyjnych w organizacjach. W modelu przełożony, na podstawie dostępności oraz postrzeganych kompetencji pracownika, przydziela członkom działu zadania. Pracownicy decydują jaki wysiłek będą wydatkować na wykonanie danej czynności. Z jednej strony na decyzje tą wpływa wiedza dotycząca własnych kompetencji, z drugiej obserwacja zachowań współpracowników. Pracownicy nie znają jednak prawdziwej czasochłonności zadania, stąd działanie nie zawsze kończy się sukcesem (wykonaniem w ustalonym czasie). Po wykonaniu kolejnych zadań przełożony otrzymuje coraz dokładniejszą informację dotyczącą kompetencji podwładnych. Jego postrzeganie nigdy jednak nie jest idealne, gdyż wykonywanie zadań stopniowo podwyższa kompetencje podwładnych w takim stopniu, w jakim pracownik wydatkował wysiłek. Natężenie obowiązków w organizacji nie jest jednostajne występują okresy, w których czas na ukończenie obowiązków jest limitowany. Zobrazowanie mechanizmu wykonywania zadań służy określeniu roli uchylania się od pracy (*shirking*) w funkcjonowaniu organizacji. Skłonność do uchylania się od pracy wydatkowania mniejszego wysiłku niż uzgodniony z pracodawcą to indywidualna charakterystyka pracownika do pewnego stopnia uzależniona od jego świadomej decyzji. Na poziomie organizacyjnym rozkład uchylania się od pracy odzwierciedla kulturę organizacyjną emergentną cechę systemu, która wypływa z ustalonych norm oraz z faktycznych działań i interakcji między członkami organizacji.

Czwartek, 29. września 2016 r., godz. 09:00 – 11:30

SESJA 2

Przewodnicząca: Jolanta Perek-Białas

Szkoła Główna Handlowa, Uniwersytet Jagielloński

Czy warto stosować Mixed Mode Data Collection? Porównanie próby zrealizowanej w eksperymencie Mixed Mode z wynikami ESS7

Paweł B. Sztabiński

Instytut Filozofii i Socjologii PAN

Obserwowany w ostatnich latach wzrost kosztów badań z zastosowaniem wywiadu osobistego przy jednoczesnym obniżaniu się ich współczynnika realizacji, kryzys badań telefonicznych oraz rozwój badań web przyczyniły się do popularyzacji Mixed Mode Data Collection. W wielu krajach, zwłaszcza zachodnich, jego stosowanie stało się normą już w początkach lat '2000. Jednak Mixed Mode oznacza w praktyce różne schematy badawcze, które mogą prowadzić w rezultacie do odmiennych wyników.

W referacie przedstawione zostaną wyniki eksperymentu Mixed Mode, zrealizowanego w Polsce paralelnie z ESS7, według tego samego kwestionariusza i na porównywalnej próbie. Zastosowano w nim połączenie schematu równoczesnego i sekwencyjnego. W pierwszej fazie zastosowano ankietę pocztową, web i EMS (kwestionariusz załączony do wiadomości e-mail), natomiast w drugiej wywiad osobisty.

Jednym z celów eksperymentu było porównanie prób zrealizowanych w Mixed Mode i w badaniu zasadniczym ESS 7, przeprowadzonym techniką wywiadu osobistego. Próby te porównano ze względu na współczynnik realizacji (*response rate*) oraz 5 cech społeczno-demograficznych. Dodatkowo uwzględniono pytania o charakterystyki respondentów, które wedle wyników badań, są powiązane z uczestnictwem w badaniach surveyowych: aktywność polityczną, zaangażowanie społeczne, orientację polityczną oraz poczucie wpływu na politykę i poczucie zagrożenia bezpieczeństwa osobistego. Biorąc pod uwagę, że respondenci mają swoje preferowane techniki badawcze, należało oczekiwać, że *response rate* będzie wyższy w Mixed Mode, a próba zrealizowana bliższa populacji, gdyż możliwość samodzielnego wypełnienia ankiety zachęci do udziału w badaniu osoby niechętnie udzielające wywiadu. Jednak rezultaty eksperymentu są niejednoznaczne. Wbrew oczekiwaniom, zastosowanie Mixed Mode nie spowodowało również obniżenia kosztów badania w porównaniu z wywiadem osobistym.

Efekty 'efektu panelowego'

Franciszek Sztabiński

Instytut Filozofii i Socjologii PAN

Mówiąc o badaniu panelowym, zawsze wspomina się o tzw. „efekcie panelowym”. Ogólnie biorąc jest to rodzaj błędu systematycznego, który pojawia się wtedy, gdy uczestnicy panelu zmieniają swoje zachowania pod wpływem

uczestnictwa w panelu. Efekt panelowy może jednak przyjmować różną postać, z zależności od tego, o jakim „panelu” mówimy.

W tradycyjnym rozumieniu, „panel” to badanie realizowane w oparciu o to samo narzędzie, w określonych odstępach czasu, na tej samej próbie jednostek badania, która jest zazwyczaj uzupełniana o młodsze kohorty wiekowe. W tym przypadku efekt panelowy może polegać na tym, że niektóre osoby uczestniczące w panelu, po pewnym czasie, zaczynają udzielać takich odpowiedzi, o których sądzą, że są właściwe, pożądane i oczekiwane przez badacza lub też, przyjmują na siebie rolę „ekspertów” (nie prezentują własnych opinii, lecz opinie „uogólnione” lub „zasłyszane”). Innym wymiarem efektu panelowego może być znużenie uczestników badania i w konsekwencji mogą oni replikować te same odpowiedzi, wybierając je w sposób rutynowy i bez zastanowienia. Należy także wspomnieć o powszechnie znanym efekcie tzw. „wymieraniu panelu”, który polega na „wykruszaniu się” badanych z panelu, a także problemach związanych z przetwarzaniem danych panelowych. Nieco inne, specyficzne, znaczenie przybiera „efekt panelowy” w przypadku badania audytorium telewizji (tzw. panel telemetryczny) i audytorium radia (tzw. panel dzienniczkowy).

Jeśli jednak mówimy o „panelu” w szerokim rozumieniu, to zazwyczaj mamy na myśli nie procedurę badawczą, lecz raczej pewnego typu operat „osób gotowych do wzięcia udziału w badaniu”, z którego losujemy jednostki mające spełniać założenia próby. Z takiego operatu (np. 90 000 internautów) losujemy jednostki do konkretnego badania, zazwyczaj ad hoc. W tym momencie dodatkowym efektem panelowym jest m.in. profesjonalizacja respondentów. Szerokie rozumienie „panelu” oczywiście nie oznacza, że na tak wylosowanej próbie nie można prowadzić badania stricte panelowego. Ale może to oznaczać multiplikację efektu panelowego.

Wpływ losowych oraz quasi-losowych procedur wewnątrzspołecznej selekcji respondentów docelowych w adresowych próbach osób na strukturę demograficzną próby oraz wskaźniki jej terenowej realizacji

Piotr Jabkowski

Instytut Socjologii UAM w Poznaniu

Badania Europejskiego Sondażu Społecznego (prowadzone na populacji osób w wieku 15 lat i więcej zamieszkujących w prywatnych gospodarstwach domowych (por. Lynn i in. 2007) opierają się na dwóch podstawowych typach prób badawczych, tj. próbach imiennych dobieranych z operatów jednostkowych oraz adresowych próbach osób (por. Sawiński 2005) dobieranych z rejestrów pozwalających na bezpośrednią lub pośrednią selekcję budynków mieszkalnych i gospodarstw domowych. O ile jednak zastosowanie operatu imiennego pozwala wprost na losowy doboru osób znanych z imienia oraz nazwiska, to wykorzystanie operatu adresowego nie daje jeszcze gwarancji uzyskania losowej próby jednostek tworzących badaną populację. W adresowych próbach osób konieczne jest bowiem zastosowanie takiej procedury wewnątrzspołecznej selekcji respondentów docelowych, która pozwoli na zobiektywizowany oraz losowy (lub quasi-losowy) dobór osób z wylosowanych wcześniej gospodarstw domowych.

W literaturze metodologicznej poświęconej badaniom surveyowym odnaleźć można opis przynajmniej kilkunastu metod służących wewnątrzspołecznemu doborowi jednostek (por. Gaziano 2005). Niektóre z takich procedur

mają charakter losowy, w tym doskonale znana siatka Kisha, inne z kolei posiadają charakter quasi-losowy, np. metody wykorzystujące daty urodzin osób wchodzących w skład gospodarstw domowych. Chociaż niepodważalną zaletą metod probabilistycznych jest — przynajmniej teoretycznie — możliwość doboru jednostek w sposób całkowicie zrandomizowany, to wymagają one jednocześnie spisania wszystkich (bez wyjątku) osób zamieszkujących wybrane gospodarstwa domowe. Prowadzi to do wydłużenia czasu potrzebnego na realizację wywiadów oraz skutkować może większymi wskaźnikami odmów, będącymi wyrazem niechęci „informerów” do ujawniania szczegółowych danych o innych domownikach zamieszkujących w wybranych gospodarstwach. Z tego względu, w procedurach quasi-losowych rezygnuje się ze spisywania wszystkich domowników, co wprawdzie skutkuje utratą pełnej losowości próby, ale jednocześnie pozwala obniżyć koszty badania oraz (zazwyczaj) prowadzi do większych wskaźników kooperacji i mniejszych wskaźników odmów.

Podstawowym celem referatu będzie porównanie dwóch najbardziej popularnych procedur wewnątrz-zespołowej selekcji respondentów (tj. siatki Kisha oraz daty urodzin). Analizy empiryczne oparte zostaną na danych pochodzących ze wszystkich siedmiu odłonek Europejskiego Sondażu Społecznego. W pierwszej kolejności autor zweryfikuje, czy rzeczywiście jest tak, że procedura daty urodzin prowadzi — zgodnie z ustaleniami literaturowymi — do większych wskaźników kooperacji oraz mniejszych wskaźników odmów. W drugiej kolejności, przeanalizowana zostanie struktura demograficzna prób badawczych opartych na selekcji respondentów docelowych poprzez siatkę Kisha oraz procedurę daty urodzin. Autor odwoła się przy tym zarówno do tzw. kryteriów zewnętrznych, polegających na porównaniu struktury demograficznej próby (z uwagi na płeć oraz wiek) ze strukturą populacji, jak również do tzw. kryteriów wewnętrznych opartych na danych uzyskanych bezpośrednio w sondażu. W tym drugim przypadku wykorzystana zostanie procedura opracowaną przez Wolfganga Sodeura (1996), polegająca na ocenie jakości określonej procedury selekcji wewnątrz-zespołowej poprzez analizę statystycznej istotności różnicy pomiędzy zanotowaną w badaniu frakcją kobiet zamieszkujących wspólnie z heteroseksualnym partnerem w dwuosobowych gospodarstwach domowych, a wzorcową — w takim typie gospodarstwach domowych frakcją kobiet równą 50%.

Słowa kluczowe: adresowe próby osób; selekcja wewnątrz-zespołowa; siatka Kisha; procedura daty urodzin

Literatura:

Gaziano, C., 2005, *Comparative Analysis of Within-Household Respondent Selection Techniques*, w: „Public Opinion Quarterly”, nr 69(1), s. 124-157.

Lynn, P., Siegfried G., Sabine H., Laaksonen S., 2007, *Methods for Achieving Equivalence of Samples in Cross-National Surveys*, w: „Journal of Official Statistics”, nr 27(1), s. 107-124.

Sawiński, Z., 2005, *Metody doboru respondentów*, s. 79-118, w: Sztabiński, Paweł B., Zbigniew Sawiński, Franciszek Sztabiński, 2005, (red.) *Fieldwork jest sztuką*, Wydawnictwo IFiS PAN, Warszawa.

Sodeur, W., 1997, *Interne Kriterien zur Beurteilung von Wahrscheinlichkeitsauswahlen*, w: „ZA-Information / Zentralarchiv für Empirische Sozialforschung”, nr 41, s. 58-82.

Refleksje na temat projektowania badań przyczynowych

Jarosław Górniak

Instytut Socjologii, Uniwersytet Jagielloński

Abstrakt zostanie zamieszczony w późniejszym terminie.

Efekt kumulacji niespójnych odpowiedzi w pytaniach zadawanych wielokrotnie na przykładzie pytań o zawód w projekcie POLPAN

Zbigniew Sawiński

Instytut Filozofii i Socjologii PAN, Instytut Badań Edukacyjnych

Wydawać by się mogło, że kolejne wywiady z tymi samymi respondentami powinny prowadzić do coraz lepszego poznania historii ich życia. Taki wniosek okazuje się jednak pochopny. Pamięć respondentów jest selektywna, przez co te same pytania zadane w różnych latach wywołują niejednakowy obraz tych samych zdarzeń. Rośnie chaos informacyjny w udzielanych odpowiedziach i po kolejnej fali wywiadów wiedza o respondentach może okazać się mniej pewna, niż była wcześniej. To paradoksalne zjawisko zilustrowane zostanie danymi na temat zawodów wykonywanych przez respondentów projektu POLPAN. Analizą objęto wyłącznie panelistów uczestniczących we wszystkich edycjach badania, z którymi wywiady przeprowadzono sześciokrotnie, w latach: 1987, 1993, 1998, 2003, 2008 i 2013 ($n = 711$). W oparciu o szczegółowe kategorie Społecznej Klasyfikacji Zawodów (SKZ) — stosowanej do kodowania we wszystkich edycjach projektu — obliczono średnią liczbę zawodów wykonywanych przez respondentów w poszczególnych latach kalendarzowych. Skalę zjawiska niespójności informacji na temat zawodu dobrze ilustrują dane dla roku 1987, ostatniego, o który pytano w pierwszej edycji POLPAN. Jeśli wziąć pod uwagę wyłącznie odpowiedzi udzielone w tym badaniu, to w 1987 roku na jednego respondenta przypadało średnio 0,92 prac. Po uwzględnieniu informacji zebranych w dalszych pięciu edycjach, średnia liczba prac dla tego roku wzrasta do 1,81. Wzrost średniej pojawia się na skutek niespójności w odpowiedziach, gdy po każdej edycji do prac podanych w poprzednich wywiadach dochodzą prace nowe, o których albo badany wcześniej nie wspomniał, albo zostały błędnie uznane przez koderów za prace różne od podanych wcześniej. Zaskakująco duża wydaje się przy tym skala wzrostu stosowanego wskaźnika. Do początkowych danych doszło aż drugie tyle informacji o zawodach niezgodnych z pierwotnie podanymi przez respondentów. Agregacja symboli SKZ w kategorii wyższego rzędu (uwzględniono podziały na 14 i 7 grup społeczno-zawodowych) w niewielkim stopniu zmniejsza liczbę niespójności, co oznacza, że biorą się one przede wszystkim z kodowania badanych do różnych klas społecznych. Jeśli w miejsce klasyfikacji posłużyć się skalami zawodów, to rozmiary niespójności ulegają zmniejszeniu, lecz ujawnia się z kolei efekt różnic w sposobach zbierania i kodowania informacji w poszczególnych edycjach badania. Wystąpienie kończącej konkluzją, że ciężar przeciwdziałania skutkom niespójnych odpowiedzi leży w poprawie jakości informacji gromadzonych w terenie. W oparciu o doświadczenia projektu POLPAN przedstawię propozycje rozwiązań w tym zakresie.

Słowa kluczowe: błędy w odpowiedziach respondentów, pomiar wielokrotny, rzetelność kodowania, badania panelowe, projekt POLPAN.

Czwartek, 29. września 2016 r., godz. 12:00 – 14:00

SESJA 3

Prowadzący: Henryk Domański

Instytut Filozofii i Socjologii PAN

Zastosowanie Magnitude scale jako narzędzia do pomiaru prestiżu społecznego. Na podstawie badań empirycznych w Polsce i Niemczech

Grzegorz Adamczyk

Instytut Socjologii, Katolicki Uniwersytet Lubelski

Peter Gostmann

Institut für Soziologie, Goethe Universität Frankfurt am Main

Sposób formułowania pytań kwestionariuszowych zaproponowany przez Renisa Likerta jest obecnie jednym z najczęściej stosowanym w badaniach społecznych. Zaletą tej formuły pytań jest oczywista — zapewnia prowadzenie pomiaru empirycznego na poziomie porządkowym, standaryzuje zmianę intensywności występowania mierzonego zjawiska i umożliwia łatwe tworzenie indeksów na podstawie baterii pytań (E. Babbie 2008, s. 197-198). Stosowanie skali Likerta wiąże się jednak z pewnymi ograniczeniami, które rozwiązuje inna skala — Magnitude scale. Pierwotnie skala Magnitude znalazła zastosowanie w empirycznych pomiarach prestiżu zawodów. Skalowanie tego rodzaju rozwinął Bernd Wegener w nawiązaniu do teorii stratyfikacji społecznej Maxa Webera. Czynnikiem wykształcania się hierarchicznego układu grup, warstw, kategorii jest społecznie „przydzielany” prestiż. „W osądach dotyczących prestiżu odzwierciedlają się społeczne granice tworzone poprzez wzajemne wykluczanie” (B. Wegener 1985, s. 229). Zdaniem Bernda Wegenera relacje pomiędzy kolektywami wyznaczającymi granice mogą być mierzone z użyciem skalowania Magnitude (B. Wegener 1985, s. 222). Metoda ta została opracowana na gruncie psychofizyki. Jej celem jest numeryczny pomiar intensywności odczuwania fizykalnych stymulantów. Procedura ta znalazła zastosowanie w badaniach preferencji politycznych (M. Lodge 1981, S. Stevens 1986, B. Wegener 1988), jak również jest proponowana dla badań dotyczących zagadnienia postrzeganego prestiżu narodowości (por. P. Gostmann, G. Wagner 2007). Podobne wnioski można sformułować w przypadku badania preferencji wartości.

Pytania o ocenę innych grup społecznych lub o ważność wartości są pytaniami o ich prestiż, tym samym stymulują respondentów do różnych, numerycznie mierzalnych reakcji. Przemawia za tym komponent emocjonalny poczucia prestiżu. Respondenci poproszeni o ocenę określonego zbioru stymulantów z użyciem dowolnych numerycznych kwantyfikatorów formułują jednocześnie osądy o relacjach pomiędzy tymi stymulantami. Rozwiązanie to ma pewne zalety w stosunku do skal porządkowych, gdy respondent używa predefiniowanych kategorii odpowiedzi. Po pierwsze, ograniczona liczba odpowiedzi (np. do pięciu lub siedmiu w przypadku skal 5-cio i 7-mio stopniowych) powoduje efekt „spłaszczania się” wyników — tzn. respondenci w rzeczywistości różnią się między sobą w ocenie danego zjawiska, ale nie mają możliwości tego zróżnicowania wyrazić — są „wtłaczani” w sztywne ramy predefiniowanej skali. Innymi słowy ci, którzy wyrażając swoje stanowisko mają do dyspozycji 5-cio stopniową

skalę, mogą nie mieć możliwości złożenia precyzyjnej deklaracji, której wyrazem byłaby np. ocena pomiędzy czwartym a piątym stopniem skali. Im krótsza predefiniowana skala, tym większa następuje utrata informacji. Po drugie, w sytuacji, gdy poszczególne stopnie skali są opisane (np. „raczej nieważne”, „raczej ważne” itd.) dokonywana przez respondenta semantyczna interpretacja ich znaczenia pozostaje niekontrolowana. Po trzecie, problematyczne jest również stosowanie skal z opisanymi wartościami najniższymi i najwyższymi. Dwaj respondenci wskazujący na skali predefiniowanej na ten sam najwyższy stopień skali, opisany np. jako „bardzo ważny”, wcale nie muszą reprezentować tej samej opinii — w rzeczywistości jeden respondent może cenić np. daną wartość o wiele wyżej niż ten drugi — tylko nie ma tego jak wyrazić (S. Schumann 2000, s. 431-432).

W kontekście powyższych uwag w referacie przedstawione zostaną wyniki badań własnych, przeprowadzonych w Niemczech i Polsce, dotyczących pomiaru prestiżu społecznego z użyciem podejścia Magnitude oraz porównawczo z użyciem „tradycyjnej” skali Likerta. Badania pretestowe w Niemczech przeprowadzono w semestrze letnim 2015/2016 wśród studentów Uniwersytetu we Frankfurcie nad Menem, badania w Polsce w czerwcu i lipcu 2016 na próbie panelu internautów Instytutu Badania Opinii i Rynku GfK Polonia.

Literatura:

Babbie Earl, *Podstawy badań społecznych*, Warszawa 2008.

Gostmann Peter, Wagner Gerhard, *Die Macht der Ehre. Eine Theorie und Methode zur Messung von Nationalprestige*, w: Peter Gostmann, Peter-Ulrich Merz-Benz (red.), *Macht und Herrschaft. Zur Revision zweier soziologischer Grundbegriffe*, Wiesbaden 2007, s. 63-79.

Schumann Siegfried, *Zur Verwendung von Magnitude-Skalen in schriftlichen Umfragen zur politischen Einstellungsforschung*, w: Jan van Deth, Hans Rattinger, Edeltraud Roller (red.), *Die Republik auf dem Weg zur Normalität? Wahlverhalten und politische Einstellungen nach acht Jahren Einheit*, Opladen 2000, s. 411-435.

Wegener Bernd, *Gibt es Sozialprestige?*, „Zeitschrift für Soziologie“ 14 (1985), s. 209-235.

Korzyści i zagrożenia związane ze stosowaniem metody IRT w międzynarodowych badaniach osiągnięć edukacyjnych

Zbigniew Sawiński

Instytut Filozofii i Socjologii PAN, Instytut Badań Edukacyjnych

W odróżnieniu od tradycyjnego podejścia do testów umiejętności, w ujęciu Item Response Theory (IRT) wynik testu nie zależy w sposób ścisły od liczby poprawnie rozwiązanych zadań, lecz oparty jest na modelu probabilistycznym. Dzięki temu metoda IRT oferuje dwie korzyści. Po pierwsze, praktycznie nieograniczoną swobodę w konfigurowaniu równoległych wersji testu, a po drugie pozwala uwzględniać informacje spoza testu, co poprawia rzetelność szacowania umiejętności. Z możliwości teorii IRT korzysta się między innymi w programie PISA, gdzie obok odpowiedzi uczniów na pytania testu bierze się pod uwagę pozostałe informacje uzyskane podczas badania, o ile wykazują związki z umiejętnościami w danej dziedzinie. Obejmuje to w pierwszym rzędzie odpowiedzi w zadaniach z pozostałych dziedzin, a ponadto wskaźniki jakości pracy

szkół, cechy położenia społecznego uczniów, czy ich zainteresowania i aspiracje. Wykorzystanie teorii IRT w programie PISA uzasadnia się potrzebą zapewnienia elastyczności w wariantowaniu zestawów zadań dla krajów z różnych obszarów kulturowych, a także potrzebą zwiększenia trafności oszacowań w warunkach niejednakowej jakości realizacji badania w różnych krajach i w różnych szkołach (błędy przy tłumaczeniu i druku zadań, brak motywacji uczniów, niewypełnianie części zadań, ściąganie i podpowiedzi). Mimo korzyści, jakich dostarcza teoria IRT, rozwiązanie to nie jest wolne od ograniczeń. Jedną z najważniejszych wad stanowi zbyt małe odseparowanie od siebie dziedzin PISA (matematyka, czytanie, rozumowanie w naukach przyrodniczych), przez co w niektórych krajach zlewają się one w jeden wynik, nie dając podstaw do oceny jakości nauczania poszczególnych przedmiotów. Problem ten zilustruję dla Polski, zestawiając wyniki PISA z wynikami centralnych egzaminów. Ponadto, niejawność procedur szacowania umiejętności wywołuje zastrzeżenia natury etycznej. Model IRT jest dopasowywany do danych za pomocą procedury iteracyjnej, przez co zasady wiążące ocenę umiejętności z konfiguracją poprawnie rozwiązanych zadań oraz z informacjami spoza testu stają się wirtualne (zilustruję to przykładem z polskich danych PISA 2012). Tymczasem w edukacji powszechny jest pogląd, że sprawiedliwa ocena powinna opierać się na zasadach jawnych i jednoznacznie rozumianych. Wiele zastrzeżeń budzi też uwzględnianie niektórych informacji spoza testu, jak chociażby pochodzenia społecznego uczniów. Wystąpienie chciałbym zakończyć dyskusją następującej kwestii: czy korzyści ze stosowania teorii IRT ważniejsze są od zagrożeń, do których prowadzi utrata kontroli nad tym, jak dane wejściowe przekładają się uzyskany wynik? Czy w międzynarodowych badaniach edukacji nie lepiej powrócić do klasycznej teorii testów, która dzięki transparentności procedur skalowania pozwala ustalić, co kryje się za różnicami w wynikach uczniów z różnych krajów.

Słowa kluczowe testy osiągnięć, Item Response Theory (IRT), klasyczna teoria testów, sprawiedliwa ocena, program PISA.

Zaufanie i satysfakcja obywateli do działania instytucji sektora publicznego w latach 2010-2014 z zastosowaniem SEM (na przykładzie Polski i Niemiec)

Magdalena Burdach, Jolanta Perek-Białas

Szkoła Główna Handlowa, Uniwersytet Jagielloński

Celem analiz jest sprawdzenie jaka jest relacja między zaufaniem, satysfakcją a oceną działania różnych instytucji publicznych na przykładzie dwóch krajów (Polski i Niemiec) w trzech okresach czasowych (2010, 2012 i 2014) na podstawie danych z Europejskiego Sondażu Społecznego (wcześniejsze analizy obejmowały tylko dwie rundy). Przed oceną modelowania weryfikowano ekwiwalentność pomiaru, tj. definicyjną, metryczną i skalarną. Następnie na podstawie zbudowanego modelu operacyjnego (w oparciu o teoretyczne wskazania) z zależnościami między oceną działalności, satysfakcją oraz zaufaniem do instytucji państwowych dokonano wyodrębnienia poszczególnych modeli pomiarowych zgodnie z konceptualizacją procesu substancywnego. Modele zbudowano dla trzech ostatnich edycji Europejskiego Sondażu Społecznego osobno oraz dla trzech rund badania łącznie, oraz oddzielnie jak i między krajami (MGCEFA). Oceny adekwatności dopasowania modeli z poszczególnymi ograniczeniami dokonano za pomocą rekomendowanych statystyk (m.in. RMSA, CFI i TLI oceniające dokładność dopasowania, oraz SRMR i CD będące miarami bezwzględnych

dopasowania modelu). Uwzględniając brak spełnienia założenia o wielowymiarowym rozkładzie normalnym badanych zmiennych jak i identyczność skali pomiarowej wszystkich wskaźników manifestowanych, oszacowanie zgodności pomiaru zoperacjonalizowanych konstruktów przeprowadzono za pomocą estymatora asymptotycznie wolnego od rozkładu (ADF). Dodatkowo estymacja modeli została przeprowadzona z zastosowaniem jak i bez odpowiednich wag, aby ustalić ich wpływ na wyniki. Modele uwzględniające wagi estymowano z zastosowaniem metody ML, tym samym celem prezentacji jest pokazanie różnic w wynikach i ocena ich istotności między metodami dostępnymi w STATA. Wyniki tych analiz będą stanowić podstawę do dyskusji na temat możliwości jak i ograniczeń stosowania modelowania równań strukturalnych w analizach tego typu zjawisk z wykorzystaniem danych ESS.

Problemy badań sondażowych: Skrajny styl odpowiadania

Katarzyna Kowalczyk, Grzegorz Król

Katedra Psychologii i Socjologii Zarządzania Wydział Zarządzania, Uniwersytet Warszawski

Style odpowiadania definiowane są jako tendencja do odpowiadania na pytania w kwestionariuszu w pewien stały sposób, niezależnie od treści samych pytań (Harzing i inni, 2011). Zjawisko to jest to szczególnie problematyczne w porównaniach międzykulturowych i może prowadzić do wyciągania błędnych wniosków np., że grupy respondentów różnią się od siebie w przedmiocie badania, podczas gdy to, co je różnicuje, to styl odpowiadania. Jak uważają niektórzy badacze style odpowiadania mogą wzmacniać lub osłabiać korelacje między wynikami respondentów na różnych skalach i wpływać na wnioski dotyczące związków między zmiennymi (Baumgartner i Steenkamp 2001). Przedmiotem naszego zainteresowania jest skrajny styl odpowiadania (ERS) definiowany jako skłonność do konsekwentnego korzystania z krańcowych odpowiedzi w kwestionariuszach z wieloma kategoriami odpowiedzi (Clark, 2000).

Literatura dotycząca ERS jest bardzo obszerna. Najczęściej pojawiając się wątki dotyczą: porównań międzykulturowych, związku ERS ze zmiennymi socjodemograficznymi i kulturowymi, metod pomiaru i problemów związanych z długością skal. Niestety wyniki tych badań często są niespójne lub niejednoznaczne (zobacz: Greenleaf, 1992; Cheung i Rensvold, 2000).

W trakcie wystąpienia zostaną zaprezentowane wyniki badań:

1. porównania stylu odpowiadania na przykładzie różnic między Japończykami i Włochami (na podstawie danych pochodzących z International Social Survey Program (1998, 2008) oraz badania internetowego reprezentatywnej próby z obu krajów)
2. wpływu ERS na wyniki badań korelacyjnych (na podstawie replikacji analiz 3 artykułów opartych na World Value Survey).

Czwartek, 29. września 2016 r., godz. 16:30 – 19:00

SESJA 4

Przewodniczący: Dariusz Przybysz

Instytut Filozofii i Socjologii PAN

Koncepcja degresywnej i progresywnej proporcjonalności i jej normatywne i opisowe zastosowania w badaniach wyborczych

Jacek Haman

Instytut Socjologii UW

„Proporcjonalność” jest pewną cechą systemu wyborczego, którą można odnosić zarówno do metody podziału mandatów, jak i do efektów działania tej metody. Wybory są proporcjonalne, jeśli do podziału mandatów wykorzystują metody dążące do rozdzielenia mandatów między partie w taki sposób, by proporcje między liczbami mandatów przyznanych poszczególnym partiom były (w przybliżeniu) takie same, jak proporcje między liczbami oddanych na nie głosów. Analogicznie, jeśli dzielimy mandaty między okręgi wyborcze (terytoria), oczekujemy zwykle, by proporcje między liczbami mandatów były takie same, jak między populacjami poszczególnych okręgów.

W wielu przypadkach jednak założeniem systemu nie jest ściśle trzymanie się proporcjonalności: systemy wyborcze (także te uznawane za „proporcjonalne”) często celowo „sprzyjają” dużym partiom (np. poprzez stosowanie progów ustawowych, lub stosowanie małych okręgów wyborczych z wysokim progiem naturalnym); zaś system podziału mandatów między terytoria w niektórych przypadkach (np. podział mandatów w Parlamencie Europejskim między państwa członkowskie UE) mogą z założenia sprzyjać podmiotom mniejszym — w takich przypadkach powinniśmy mówić nie tyle o „proporcjonalności”, co o, odpowiednio, „progresywnej” i „degresywnej proporcjonalności”.

Pojęcie „degresywnej proporcjonalności” od lat funkcjonuje w kontekście dyskusji o zasadach podziału mandatów w PE — z reguły jednak nie była mu nadawana ścisła treść. Propozycję ścisłej definicji „degresywnej proporcjonalności” oraz odwołującą się do niej metodę podziału mandatów zaproponowałem już kilka lat temu. W prezentowanym referacie, po przypomnieniu tej propozycji, przedstawię jej uogólnienie obejmujące również „progresywną proporcjonalność”, a także przedstawię możliwość zastosowania tych koncepcji nie tylko do rozwiązywania kwestii normatywnych (jak dzielić mandaty), ale również do celów deskryptywnych: jak mierzyć własności systemów wyborczych, a ściślej — jak mierzyć, na ile dany podział mandatów jest „skrzywiony” na korzyść dużych (ewentualnie małych) partii. Przedstawię również przykłady analiz wykorzystujące opisywaną koncepcję.

Nierówność i jej postrzeganie: W poszukiwaniu źródeł wielowymiarowości

Barbara Jancewicz

Instytut Socjologii UW

Badacze starają się powiązać nierówność z wieloma negatywnymi zjawiskami społecznymi takimi jak przestępczość, wzajemna nieufność czy nawet zwiększona umieralność (Wilkinson i Pikett, 2011). Poziom nierówności ocenia się przy pomocy specjalnie skonstruowanych miar nierówności: każda z nich odzwierciedla jedną z wielu funkcjonujących koncepcji tego pojęcia. Ekonomisci nie są bowiem jednomyślni co do tego, czym dokładnie jest nierówność, więc równolegle stosują kilka różnych definicji. Amiel i Cowell (1999) postanowili sprawdzić, na ile definicje stosowane przez ekspertów odpowiadają tym używanym przez zwykłych ludzi: zapoczątkowali oni serię badań nad ocenami nierówności dochodowej.

Badania ocen nierówności dochodów zakładają, że każdy respondent jest racjonalny i posiada jednowymiarowe wyobrażenie na temat nierówności. Czasem wydobycie tego wyobrażenia wymaga namysłu, ale oczekuje się od badanych, że w końcu udzielą jednowymiarowej i spójnej odpowiedzi. Wyniki dotychczasowych badań pokazały, że rzeczywistość jest dużo bardziej skomplikowana, a odpowiedzi respondentów często niejasne. Inspirując się uwagą Amartyi Sena (1973) wskazującego, że nierówność ma wiele aspektów przeprowadziłam badanie, które pokazało, że oceny respondentów najlepiej odzwierciedlała dwuwymiarowa mapa. Niniejsze badanie kontynuuje analizę tego problemu: testuje, czy dwuwymiarowość jest efektem różnic między badanymi, czy też pojawia się ona już w ocenach pojedynczych badanych.

Sprawdzenie, czy poszczególne osoby postrzegają nierówność wielowymiarowo wymagało stworzenia nietypowego kwestionariusza. Indywidualne odpowiedzi zobrazowano przy pomocy skalowania wielowymiarowego na jednym, dwóch oraz trzech wymiarach. Analiza poziomu błędów dopasowania pozwoliła na stwierdzenie, że indywidualne odpowiedzi najlepiej odzwierciedla pojedyncza oś: czyli zgodnie z dotychczas przyjmowanymi założeniami. Odpowiedź każdego badanego została odzwierciedlona na indywidualnej osi, ale osie różnych badanych były różne. Zobrazowanie odpowiedzi wielu badanych wymagało więc zastosowania większej liczby wymiarów.

Nowatorski kwestionariusz i nietypowe zastosowanie skalowania wielowymiarowego stanowią o innowacyjności metodologicznej prezentowanego badania. Sformułowane wnioski pozwolą lepiej badać, a w konsekwencji lepiej rozumieć, w jaki sposób ludzie oceniają nierówności.

Polaryzacja ekonomiczna w Polsce

Tomasz Panek

Szkoła Główna Handlowa

W referacie zostanie przedstawiona ocena czy prowadzona po akcesji Polski do Unii Europejskiej polityka społeczna i gospodarcza doprowadziła do zmniejszenia dysproporcji dochodowych gospodarstw domowych w Polsce. Zmiany dysproporcji dochodowych gospodarstw domowych były analizowane poprzez ocenę zmian stopnia polaryzacji ekonomicznej. Ponadto jako uzupełnienie analizy procesu polaryzacji ocenie poddano zmiany nierówności dochodowych. Ana-

lizy zmian stopnia polaryzacji dochodowej oraz nierówności dochodowych gospodarstw domowych w Polsce zostały przeprowadzone dla lat 2000-2014 (zarówno przed akcesją jak i po akcesji Polski do Unii Europejskiej). W analizie procesu polaryzacji zostały zastosowane indeksy polaryzacji, a zmian nierówności dochodowych współczynniki nierówności.

Miara różnorodności Giniego-Simpsona

Grzegorz Lissowski

Instytut Socjologii, Uniwersytet Warszawski

Miara różnorodności, nazywana tak od nazwisk statystyków, którzy ją proponowali C. Giniego (1912) i E.H. Simpsona (1949), a niekiedy także różnorodnością klasyfikacji, będzie przedstawiona na tle klasy parametrów różnorodności. Przedstawione będą:

- własności tej miary,
- jej interpretacja oraz
- aksjomatyczna charakterystyka.

Miara ta i jej własności będą porównane przede wszystkim z entropią. W zakończeniu pokazane zostaną relacje między parametrami różnorodności a parametrami koncentracji.

Międzyosobowa zgodność odpowiedzi w badaniach ankietowych: Zróżnicowanie poglądów w perspektywie wewnątrz i międzykrajowej

Tadeusz Krauze

Hofstra University

Przemek Powąłko

Instytut Filozofii i Socjologii PAN

Kazimierz M. Słomczyński

Instytut Filozofii i Socjologii PAN, The Ohio State University

Rezultaty badania sondażowego można przedstawić w tabeli x_{ik} wymiaru m na n , gdzie i oznacza numer respondenta $1 \leq i \leq m$, k oznacza numer pytania $1 \leq k \leq n$, x_{ik} oznacza odpowiedź i -tego respondenta na pytanie k . W sondażu wyróżniamy trzy rodzaje zmiennych (czyli pytań): zmienne techniczne, zmienne dotyczące faktów, oraz zmienne (pytania), na które odpowiedź jest subiektywna, wyrażająca postawy, opinie, wartości i preferencje. Tylko ten trzeci rodzaj zmiennych jest indeksowany przez symbol k .

Poglądem i -tego respondenta nazywamy ciąg odpowiedzi subiektywnych, czyli i -ty wiersz tabeli: $x_{i1}, x_{i2}, \dots, x_{in}$. Następnie porównujemy poglądy dwóch dowolnych respondentów i, j . Jeżeli odpowiedzi respondentów i, j na k -te pytanie są takie same (czyli $x_{ik} = x_{jk}$), to przypisujemy tej różnicy wartość 0; jeżeli $x_{ik} \neq x_{jk}$ to przypisujemy wartość 1. Suma jedynek jest miarą różnicy poglądów między dwoma dowolnymi respondentami. Jest to odległość Hamminga $d(i, j)$:

$$d_{ij} = \sum \text{ind}(x_{ik} - x_{jk}),$$

gdzie $\text{ind}(z) = 0$, gdy $z = 0$, $\text{ind}(z) = 1$, gdy $z \neq 0$. Mamy $0 \leq d_{ij} \leq n$, ponieważ jest n subiektywnych odpowiedzi. Jeżeli $d_{ij} = 0$, to wszystkie odpowiedzi

respondentów i, j są jednakowe; jeżeli $d_{ij} = n$, to odpowiedź respondenta i na każde pytanie jest inna niż odpowiedź respondenta j . Wartość d_{ij}/n jest proporcją niezgodnych odpowiedzi respondentów i, j . Maksymalna liczba par różnych respondentów jest $m(m-1)/2$. Indeks D jest globalną miarą niezgodności poglądów:

$$D = \frac{2}{m(m-1)} \sum_{i=1}^m \sum_{j>i} \frac{1}{n} d_{ij}$$

Indeks D jest średnią proporcji niezgodnych odpowiedzi we wszystkich parach i, j gdzie $j > i$. Indeks D jest unormowany: $0 \leq D \leq 1$. Indeks D może być użyty do porównywania sondaży, ponieważ jest niezależny od liczby respondentów i liczby pytań.

Dane empiryczne, które mogą być analizowane za pomocą indeksu D , pochodzą z badania nad harmonizacją danych sondażowych. W projekcie tym jest 1721 narodowych sondaży pokrywających 142 kraje/terytoria, w których brało udział ponad 2,2 miliona respondentów. Ponieważ próby respondentów były w zamierzeniu reprezentatywne, można powiedzieć, że odzwierciedlają one społeczeństwa.

Przyпускаjmy, że wartość D jest stabilna w czasie dla każdego kraju; można to sprawdzić, ponieważ w 27 krajach przeprowadzono po 20 i więcej sondaży. Przyпускаjmy, że D różni się znacznie pomiędzy krajami, ale nie różni się w podzbiorach pytań (tak zwanych blokach tematycznych). Analiza takich różnic wymaga dekompozycji indeksu D . W przypadku s bloków, mamy $n = n_1 + n_2 + \dots + n_s$; w dekompozycji używamy średniej ważonej liczb n_1, n_2, \dots, n_s .

Przedstawione podstawy metodologiczne projektu badawczego skoncentrowane są na zróżnicowaniu poglądów (*dissensus* jako opozycja do *consensus*), czyli na strukturze opinii publicznej. Klasycy socjologii — Simmel, Tönnies, Weber, Marks, Toqueville, Durkheim — żeby wymienić najbardziej znaczących przedstawicieli, zajmowali się tymi pojęciami, często *implicite* i zawsze w ujęciu jakościowym. Nasze podejście jest ilościowe i zostanie wykorzystane w dalszej pracy nad zróżnicowaniem poglądów w perspektywie wewnątrz- i międzykrajowej.

Piątek, 30. września 2016 r., godz. 09:00 – 12:30

SESJA 5

Przewodniczący: Zbigniew Sawiński

Instytut Filozofii i Socjologii PAN, Instytut Badań Edukacyjnych

Zastosowanie schematu analizy *difference-in-differences* w badaniach politycznych

Adam Gendźwiłł

Wydział Geografii i Studiów Regionalnych, Uniwersytet Warszawski

Tomasz Żółtak

Metody quasi-eksperymentalne zyskują obecnie coraz większą popularność, zarówno w analizach zorientowanych bardzo praktycznie, jak ewaluacja wpływu programów w edukacji czy polityce społecznej, jak również w badaniach podstawowych. Pozwalają one w możliwie wiarygodny sposób wnioskować o wielkości oddziaływania interwencji (zdarzeń), w szczególności wszelkiego rodzaju reform, w sytuacjach, gdy prowadzenie badań w schemacie eksperymentalnym jest niemożliwe lub nieopłacalne. W trakcie wystąpienia omówiona zostanie jedna z szerokiego i zróżnicowanego spektrum metod quasi-eksperymentalnych, jaką stanowi schemat analizy *difference-in-differences*, ze szczególnym uwzględnieniem jego zastosowania w badaniach politycznych.

W pierwszej części wystąpienia metoda *difference-in-differences* scharakteryzowana zostanie pod względem formalnym. Jej założenia omówione zostaną w odniesieniu do zaproponowanej przez Donalda Rubina ramy teoretycznej *potential outcomes* (określanej też jako model przyczynowy Neymana-Rubina), pozwalającej zintegrować opis zasad wnioskowania przyczynowego w badaniach z użyciem schematów eksperymentalnych i quasi-eksperymentalnych. Następnie przedstawione zostaną możliwości modyfikacji podstawowego schematu *difference-in-differences* oraz wykorzystywane w takich przypadkach metody estymacji efektu interwencji. Omówione zostaną też ograniczenia metody, a w szczególności kwestia konieczności przetestowania, czy spełnione są założenia przyjmowane w modelu.

Druga część wystąpienia będzie mieć charakter aplikacyjny. Na początku możliwości stosowania schematu *difference-in-differences* zostaną krótko porównane z możliwościami wykorzystania innych szeroko wykorzystywanych metod quasi-eksperymentalnych: *propensity score matching*, *inverse probability of treatment weighting*, *regression discontinuity design* i *synthetic control group*. Następnie przedstawione zostaną przykłady praktycznego zastosowania metody w badaniach z zakresu socjologii polityki. Obejmą one analizy skutków ostatnich reform polskiego prawa wyborczego: zmiany systemu wyborczego w wyborach do rad niektórych gmin w wyborach samorządowych 2014 r. oraz skutków wprowadzenia zbroszurowanych kart do głosowania.

Problemy pomiaru efektów kontekstowych bieżących wydarzeń w międzykrajowych badaniach porównawczych. Wyniki badań ESS Runda 6 i 7

Teresa Żmijewska-Jędrzejczyk
Instytut Filozofii i Socjologii PAN

Jednym z wymiarów oceny wartości danych z międzykrajowych badań porównawczych jest homogeniczność, zapewnienie standaryzacji sytuacji badawczej. Do jej nieodłącznych elementów należy kontekst bieżących wydarzeń, którego wystandaryzować się nie da, a kontrolować go jest trudno. Skrajne przykłady uwzględniania kontekstu w procesie badawczym odnajdujemy w literaturze (Vyrost J., Fedakova D., Kentos M. 2007, Stathopoulou T. 2007). Polegały one na zawieszeniu realizacji terenowej na poziomie kraju, w celu utrzymania homogeniczności sytuacji badawczej i możliwości włączenia danych do zbioru międzynarodowego. Uwzględnianie efektów kontekstowych ujawniających zdestandaryzowanie sytuacji badawczej jest do pewnego stopnia możliwe także po zakończeniu realizacji terenowej, na etapie oceny homogeniczności wyników badania. Zilustruję to na przykładzie ESS7 we Francji, sprawdzając czy bieżące wydarzenie może wpływać na zmiany opinii i postaw.

Mimo korzyści płynących z uwzględniania kontekstu bieżących wydarzeń, metody stosowane do jego monitorowania mają pewne specyficzne ograniczenia. ESS jest jedynym badaniem przekrojowym realizowanym w różnych krajach, w którym bieżące wydarzenia są rejestrowane. Z tego względu przedstawione analizy dotyczą tylko tego badania. W ESS monitoring kontekstu jest realizowany wg Claims Making Approach (Statham P., Tumber H. 2009). W metodzie tej przyjmuje się, że jedynie wydarzenia bezpośrednio powiązane tematycznie z treścią kwestionariusza mogą wpływać na proces formułowania odpowiedzi w danej kwestii. Monitoring wydarzeń obejmuje dzienniki ogólnoinformacyjne, z wyłączeniem tabloidów. Weryfikacja tych założeń stanowiła jeden z celów badania zrealizowanego po zakończeniu rundy ESS rundy 6. Wyniki tego badania oraz ESS7 dla Polski przedstawię w drugiej części referatu.

Konkluzje z przeprowadzonych badań są następujące. Śledzenie efektów kontekstowych bieżących wydarzeń jest istotnym elementem oceny ich wartości pod względem homogeniczności. Pomiar tych efektów stanowi nadal wyzwanie dla badaczy a obecnie stosowane metody są niewystarczające. Monitorowanie wydarzeń wyłącznie na podstawie przeglądu dzienników ogólnoinformacyjnych (z wyłączeniem tabloidów) oraz rejestrowanie tylko tych wydarzeń, które są literalnie powiązane tematycznie z treścią pytania stanowią najpoważniejsze ograniczenia metody. Przeprowadzone analizy ujawniły, że wydarzenia niezwiązane tematycznie, nieobecne w dziennikach ogólnoinformacyjnych wpływają na proces formułowania opinii w odniesieniu do pytań o zaufanie do instytucji publicznych i ich ocenę.

Synthetic Control jako metoda badania skutków reform terytorialnych

Paweł Swianiewicz, Julita Łukomska
Zakład Rozwoju i Polityki Lokalnej, Wydział Geografii i Studiów Regionalnych, Uniwersytet Warszawski

Dotychczasowe prace prowadzą do wniosku, że bezpośredni pomiar zjawiska korzyści skali w odniesieniu do usług wykonywanych przez samorządy jest

metodycznie niezmiernie skomplikowany. Próba odpowiedzi na te trudności jest potraktowanie reform terytorialnych jako laboratorium badawczego. Reformy podziału terytorialnego stwarzają w tym kontekście unikalne warunki. Możemy je traktować jako rzadko dostępne w naukach społecznych laboratoria umożliwiające obserwację wpływu samej zmiany wartości zmiennej wyjaśniającej (w naszym przypadku wielkości jednostki samorządowej) na wartość zmiennej zależnej (w naszym przypadku kosztów wykonywanych usług), nieomal bez zmian wartości pozostałych zmiennych interweniujących.

Przedmiotem naszej analizy empirycznej są efekty zmian terytorialnych w Polsce:

- w 2002 roku powstało 7 nowych powiatów, wyodrębnionych z funkcjonujących wcześniej większych jednostek;
- w okresie 1995-2010 miało miejsce powstanie 15 nowych gmin (najnowszym takim przypadkiem było oddzielenie się gminy Jaślika od gminy Dukla w woj. podkarpackim; pozostałe nowe gminy tworzone były jeszcze w ostatnich latach XX wieku).

Sytuację w podzielonych gminach i powiatach można porównywać z sytuacją w podobnych jednostkach, które nie podlegały w tym czasie zmianom terytorialnym – w ten sposób można przeprowadzić badanie w schemacie quasi-eksperymentalnym, który wzmacnia argumentację o zależności przyczynowej między wielkością gminy a jej sytuacją finansową. W naszym badaniu przyglądamy się więc nowo powstałym jednostkom analizując zmiany wartości dwóch zmiennych odnoszących się do kosztów wykonywania zadań i sytuacji finansowej samorządów: wydatków administracyjnych i wielkości nadwyżki operacyjnej budżetu. Analizujemy przy tym nie tylko wydatki bezpośrednio przed i po przeprowadzonej zmianie granic, ale przyglądamy się kilkuletnim trendom. Objęcie obserwacją kilku lat wstecz przed reformą uzasadnione jest obserwacjami Blom-Hansena (2009) i oraz Kjæra et al. (2010) o tym, że zachowanie samorządów bezpośrednio przed spodziewaną zmianą jest często nietypowe.

Zastosowaliśmy metodę *Synthetic Control* (Abadie et al. 2010), której istotą jest stworzenie dla każdego z badanych przypadków abstrakcyjnej „jednostki kontrolnej” – abstrakcyjnego konstruktów zbudowanego w oparciu o ważne wartości badanych wskaźników w kilku podobnych jednostkach (odpowiednio gminach lub powiatach). Im podobniejsze wartości zmiennych do badanego przypadku w okresie przed podziałem tym wyższa jest ich waga w konstrukcji „jednostki kontrolnej”. Kryteria podobieństwa odnoszą się do analizowanych zmiennych zależnych (wydatków na administrację oraz wielkości nadwyżki operacyjnej) przed podziałem, a także innych czynników, które mogą mieć wpływ na trendy zmian tych zmiennych (liczba ludności, gęstość zaludnienia, zamożność społeczności lokalnej – mierzona wielkością lokalnej bazy podatkowej – oraz położenie w ramach tego samego regionu NUTS-1). Następnie sprawdzamy czy trend zmian w kolejnych latach wykazuje istotne różnice między jednostką badaną, a „jednostką kontrolną”. W efekcie pozwala to na zidentyfikowanie wpływu zmiany wielkości jednostki samorządowej, przy założeniu, że inne czynniki (takie jak zamożność, gęstość zaludnienia itp.) pozostają takie same w obu porównywanych jednostkach.

Metody międzykrajowych badań sondażowych nad post-materializmem i post-modernizmem (szkoła R. Ingleharta): krytyka i propozycje własne

Andrzej J. Kutyłowski

Uniwersytet w Oslo

Pierwsza część prezentacji zawierać będzie skrótowe wprowadzenie do tezy o „post-materializmie” (R. Inglehart i inni) jako „zmiany wartości”, z krótkim zarysem historii tej tezy oraz jej poznawczej roli jako „typu idealnego”, skonstruowanego z pożądanym „typem realnym”. Uwzględnione też zostanie rozszerzenie tezy o „post-materializmie” do tezy o „post-modernizmie”, rozwiniętej przez R. Ingleharta w oparciu o dane z europejskich i światowych sondaży nt. „wartości” (np. World Value Surveys) i stanowiące przedmiot aktualnych badań szeregu autorów.

Druga część wystąpienia dotyczyć będzie metod pomiaru „post-materializmu”, w kontekście nadania im interpretacji stochastycznej, z odniesieniem do metody uzyskiwania danych jako realizacji pewnego schematu preferencyjnego oraz wyszczególnienia metod statystycznych uwzględniających specyfikę pomiaru „post-materializmu” w badaniach międzykrajowych. Zaproponowane zostaną pewne nowe metody analizy, które nie występują w literaturze przedmiotu. Wspomnę też o możliwościach rozszerzenia schematu preferencyjnego Ingleharta.

Trzecia i ostatnia część wystąpienia poświęcona zostanie krytyce metod stosowanych przez Ingleharta i inn. w analizach post-materializmu i post-modernizmu oraz tym podobnych. Zamiast tradycyjnych metod opartych na błędnym (aczkolwiek z reguły nie wypowiedzianym) założeniu, że wskaźniki kategoriowe (*categorical*) mogą być traktowane jako wskaźniki metryczne (*metrical*), oraz dodatkowym i z reguły błędnym założeniu, że pseudo-metryczne wskaźniki są rozłożone wg. rozkładu Gaussa, przedstawione zostaną alternatywne ramy modelowania, oparte o idee probabilistycznych modeli graficznych dla zmiennych kategoriowych.

Bardziej szczegółowe streszczenie z 13 odnośnikami dostępne jest w języku angielskim od autora lub organizatorów konferencji.

Wykorzystanie metod obliczeniowych i Pythona w analizach danych

Grzegorz Król

Katedra Psychologii i Socjologii Zarządzania, Wydział Zarządzania Uniwersytet Warszawski

Klasyczne metody statystyczne są przez wielu naukowców spostrzegane jako trudne. Czasami tak trudne, że skutecznie zniechęcają badaczy do stosowania metod ilościowych. W innych sytuacjach sprawiają, że naukowcy bezrefleksyjnie stosują popularne testy statystyczne, które — bez weryfikacji założeń dotyczących np. rozkładu zmiennych czy równości wariancji w grupach — mogą generować niepoprawne wyniki i skutkować błędnymi interpretacjami rezultatów badań empirycznych.

Dramatyczny postęp szybkości i mocy obliczeniowej komputerów, dostępnych obecnie dla każdego badacza, umożliwia zmianę podejścia, z analitycznego, opartego na szeregu założeń teoretycznych, na podejście obliczeniowe, w którym część „klasycznych” założeń staje się zbędna, a często także liczba

kroków wymaganych do przetestowania hipotez zostaje zredukowana, upraszczając proces analizy danych, a w rezultacie ułatwiając zrozumienie poszczególnych etapów tego procesu, i poprawną interpretację wyników.

W prezentacji przedstawione zostaną, w przystępny sposób, przykłady wykorzystania języka programowania Python do rozwiązania przykładowych problemów analizy danych w naukach społecznych i zarządzaniu. Możliwości uproszczenia analiz w podejściu obliczeniowym podkreślone będą poprzez ograniczenie w większości prezentowanych przykładów kodu języka Python do zaledwie 4 lub 5 wierszy, wymaganych do rozwiązania każdego prezentowanego problemu.

Jako alternatywna dla weryfikacji hipotez przy użyciu znanych rozkładów statystyk (t , chi-kwadrat, itp.) wykorzystane zostaną takie techniki obliczeniowe jak symulacja, tasowanie, bootstrapping, czy walidacja krzyżowa.

W prezentacji uwypuklone zostaną także walory dydaktyczne wykorzystania przedstawionych metod w uczeniu klasycznych pojęć statystycznych jak „hipoteza zerowa”, „wartość p ” czy „rozkład z próby”.

Prezentację zakończy wskazanie warunków, w których klasyczne podejście do analiz statystycznych pozostaje wciąż skuteczniejsze niż prezentowane podejście obliczeniowe.