

Metodologiczne Inspiracje 2023

Badania Ilościowe w Naukach Społecznych — Wyzwania i Problemy

21-22 września 2023, Jabłonna

21 września

Sesja 1. Prowadzący: Dariusz Przybysz (IFiS PAN)

ZBIGNIEW SAWIŃSKI, INSTYTUT FILOZOFII I SOCJOLOGII PAN

ANNA KIERSZTYN, UNIWERSYTET WARSZAWSKI

Ogólny schemat harmonizacji danych z badań panelowych

Harmonizacja danych biograficznych pochodzących z badań panelowych wymaga uzupełnienia dotychczasowych schematów harmonizacji o narzędzia uwzględniające specyficzne własności i rozwiązania stosowane w badaniach panelowych w różnych krajach. W prezentacji proponujemy schemat, uwzględniający, obok innych charakterystyk badań wpływających na porównywalność danych, dwie specyficzne cechy badań panelowych. Po pierwsze, dane pochodzące z różnych fal badania nie zawsze są spójne. Po drugie, dla niektórych okresów indywidualnych karier dostępne mogą być dane z więcej niż jednej fali.

Nasze podejście zakłada przekształcenie oryginalnych zbiorów danych panelowych do ujednoliconego formatu i jednocześnie uzupełnia te dane o kody ASC (*Activity Status Codes*) odzwierciedlające kompletność i sposób uzyskania danych dotyczących aktywności (pracy zawodowej, nauki, przebywania na urlopie itd.) podejmowanych przez respondentów w kolejnych jednostkach czasu. Klasyfikacja ASC pozwala w szczególności na rozróżnienie informacji o dacie rozpoczęcia i zakończenia danej aktywności podanych przez respondenta lub imputowanych podczas edycji danych, bądź też wynikających z organizacji badania. To ostatnie dotyczy sytuacji, w której rozpoczęcie bądź zakończenie aktywności jest skutkiem odcięcia części danych przez moment realizacji wywiadu (*censoring*) oraz sposób zadawania pytań, jeśli pytania obejmują ograniczony przedział czasowy (np. praca tylko po ukończeniu szkoły, luki związane z wymieraniem panelu bądź przerwami panelistów w uczestnictwie w badaniu). Obok kodów ASC, do przekształconego zbioru dodawane są zmienne metodologiczne, w tym informacje dotyczące fali badania, z której pochodzą dane o poszczególnych aktywnościach. Pozwala to użytkownikom danych na łatwe porównania danych opisujących różne okresy biografii respondentów zarówno między badaniami panelowymi, jak i między różnymi falami tego samego badania. Zostawia zarazem swobodę wyboru między wykorzystaniem jedynie danych z najbliższej fali bądź uwzględnieniem

również danych z późniejszych fal panelu. Ten ostatni wybór zasadny jest szczególnie wtedy, gdy dane z jednej z fal zawierają odmowy, bądź cechują się niższą jakością.

Proponowane podejście zilustrowane zostało porównaniem danych na temat historii zatrudnienia, zebranych w polskim badaniu panelowym POLPAN oraz niemieckim panelu społeczno-ekonomicznym G-SOEP. Oba badania znacząco różnią się metodologią. POLPAN realizowany jest w cyklu pięcioletnim, przy czym o każdą z prac pyta się osobno, ustalając datę jej rozpoczęcia i zakończenia. W panelu G-SOEP dane zbiera się co roku, zaś pytania na temat rodzajów aktywności, w tym wykonywanie prac, zadawane są w formie kalendarza. W prezentowanym przykładzie dane w obu badaniach przekształcone zostały do rekordów rocznych, a następnie przypisano im symbole ASC i uzupełniono o zmienne metodologiczne. Przykład ilustruje sposób, w jaki proponowane podejście ułatwia identyfikację i oszacowanie potencjalnych konsekwencji odmiennej metodologii badań, tak aby w ostatecznym bilansie możliwe było uwzględnienie tych różnic w analizach porównujących przebieg karier zawodowych w obu krajach.

Choć nasze podejście wypracowane zostało dla celów porównania danych o karierach zawodowych młodych dorosłych w czterech krajach analizowanych w ramach projektu Cross-National Biographies – Young (<https://crossnationalbiographies.edu.pl/>), skrypty obejmujące konwersję danych i kodowanie w systemie ASC mogą być upowszechniane, modyfikowane i stosowane w innych projektach. Tym samym omawiane podejście otwiera nową platformę gromadzenia rozwiązań pomocnych w harmonizacji badań panelowych z różnych krajów.

Kluczowe terminy: badania panelowe, harmonizacja danych, historie zatrudnienia, badanie POLPAN, badanie G-SOEP

MARIA FLAKUS, INSTYTUT FILOZOFII I SOCJOLOGII PAN

Nauka w czasach post-API – (aktualnie) dostępne sposoby pozyskiwanie danych do celów naukowych z portalu Facebook

Wzrost liczby platform społecznościowych (*social media sites*, SMS) i zwiększenie ich różnorodności stanowią jedne z najważniejszych zjawisk obecnych w przestrzeni internetowej w ostatnich latach. Pomimo mnogości SMS oraz upływu lat, platformą o największej liczbie użytkowników pozostaje Facebook, będący przedmiotem zainteresowania różnych grup naukowców, podejmujących zarówno badania kontekstu użytkownika Facebooka (np. motywacji czy też zachowań towarzyszących ekspozycji na to medium), jak i próby wykorzystywania danych o aktywności użytkowników do modelowania cech niedostępnych w SMS *explicite* (np. cech osobowości lub poglądów politycznych). W odniesieniu do drugiej z wymienionych przestrzeni problemowych, w przeszłości badacze wykorzystywali dane z samej platformy, gromadzone za pośrednictwem jej interfejsów programowania aplikacji (Application Programming Interfaces, API). Jakkolwiek informacje gromadzone w ten sposób pozwoliły na sformułowanie ciekawych wniosków i rozwój wielu obszarów nauk społecznych (np. psychologii społecznej czy osobowości), to dostęp do danych gromadzonych w ten sposób został znacząco ograniczony po skandalu z udziałem firmy Cambridge Analytica, stając się zasadniczo niedostępnym dla naukowców niezwiązanych z korporacją Meta. Mając na uwadze bogactwo danych, jakie potencjalnie związane są z użytkowaniem SMS, wielu badaczy zaczęło akcentować konieczność wypracowania nowych strategii uzyskiwania dostępu do danych z platformy, pośród których wymienia się metody takie jak: wtyczki obserwujące aktywność w mediach społecznościowych, bezpośrednia współpraca z użytkownikami platformy w ramach grup lub stron stworzonych w

platformie oraz pozyskiwanie danych bezpośrednio od uczestników. W niniejszym wystąpieniu omówione zostaną te i inne sposoby pozyskiwania danych z Facebooka, eksponując konteksty metodologiczne (wady i zalety poszczególnych podejść) i etyczne (m.in. w kontekście zagrożeń dla osób badanych) związane z każdą z metod.

PIOTR CICHOCKI, WYDZIAŁ SOCJOLOGII, UAM W POZNANIU

PIOTR JABKOWSKI, WYDZIAŁ SOCJOLOGII, UAM W POZNANIU

Analiza wskaźników realizacji prób sondażowych w czterech międzykrajowych projektach porównawczych realizowanych w Europie w ostatnich 40 latach – ustalenia empiryczne

Prezentacja opiera się na danych zgromadzonych w oparciu o analizę dokumentacji metodologicznej czterech dużych projektów porównawczych o charakterze sondażowym realizowanych w Europie od 1981 roku: European Quality of Life Survey (EQLS), European Social Survey (ESS), European Values Study (EVS) oraz International Social Survey Programme (ISSP). W zaprezentowanej analizie wykorzystaliśmy dane dotyczące implementacji badań terenowych i charakterystyk metodologicznych ponad 1,4 tysiąca krajowych. Celem tej analizy było ustalenie relacji pomiędzy trzema wskaźnikami realizacji prób sondażowych (*response rate*, *refusal rate*, *contact rate*) a wykorzystywanymi procedurami badawczymi. Wyniki naszych analiz potwierdzają ustalenia innych badaczy a mianowicie, że wskaźniki realizacji badania maleją w czasie. Spadek jest szczególnie widoczny w EVS, w którym wskaźnik *response rate* spadł z około 65% w 1999 roku do 43% w najnowszej rundzie datowanej na 2017 rok. Podobną skalę spadków zaobserwować można również w ISSP oraz EQLS, z dużo mniejszym efektem czasu obserwowanym w ESS. Nasze ustalenia pokazują dodatkowo, że spadek wskaźników realizacji prób sondażowych jest niezależny od typu próby badawczej, zastosowanego schematu losowania oraz wdrożonych procedur terenowych. Wyniki sugerują zatem, że trudności w realizacji prób sondażowych są nieuniknione. Ten wniosek ma istotne konsekwencje dla praktyki badawczej, ponieważ oznacza, iż realizacja sondaży wymagać będzie losowania prób o większej liczebności oraz dużo większych nakładów finansowych związanych z nakłanianiem respondentów do udziału w badaniu.

MAREK BOŻYKOWSKI, OŚRODEK PRZETWARZANIA INFORMACJI – PAŃSTWOWY INSTYTUT BADAWCZY/UNIwersytet Warszawski

AGNIESZKA CHŁOŃ-DOMIŃCZAK, OŚRODEK PRZETWARZANIA INFORMACJI – PAŃSTWOWY INSTYTUT BADAWCZY/ SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE

MIKOŁAJ JASIŃSKI, OŚRODEK PRZETWARZANIA INFORMACJI – PAŃSTWOWY INSTYTUT BADAWCZY/UNIwersytet Warszawski

PRZEMYSŁAW KOPA, OŚRODEK PRZETWARZANIA INFORMACJI – PAŃSTWOWY INSTYTUT BADAWCZY/UNIwersytet Warszawski

Sekwencje wchodzenia na rynek pracy absolwentów szkół wyższych

Dane administracyjne pozwalają na prowadzenie wzdłużnych analiz badanych zjawisk. Do takich należy analiza sekwencji statusów na rynku pracy. Dane pochodzące z Ogólnopolskiego systemu monitorowania ekonomicznych losów absolwentów szkół wyższych (ELA) pozwalają na analizę zjawisk edukacyjnych i zawodowych z dynamiką miesięczną.

W naszej prezentacji przedstawimy analizę uwarunkowań sekwencji wejścia na rynek pracy absolwentów studiów magisterskich różnych roczników - w tym absolwentów wychodzących na rynek pracy w okresie pandemii COVID-19. Różne sekwencje wejścia wiążą się z różnymi charakterystykami pozycji na rynku pracy (w tym zarobków).

Do wyznaczenia macierzy odległości (niepodobieństw) pomiędzy poszczególnymi sekwencjami statusu edukacyjno-zawodowego wykorzystaliśmy metodę *optimal matching*. Następnie przy użyciu hierarchicznej analizy skupień opracowaliśmy systematykę sekwencji statusu. W prezentacji wskażemy, jakie cechy badanych sprzyjają wejściu na korzystne i niekorzystne ścieżki na rynku pracy. W analizie zostały uwzględnione indywidualne cechy badanych, cechy ukończonych przez nich studiów oraz charakterystyki ich powiatów zamieszkania. Rozważania uzupełnione będą odniesieniem uzyskanych wyników do kalendarium pandemii i lockdownu.

SESJA 2. Prowadzący: Piotr Jabkowski (Wydział Socjologii, UAM w Poznaniu)

ARTUR POKROPEK, INSTYTUT FILOZOFII I SOCJOLOGII PAN

Identyfikacja nieuważnego odpowiadania w ankietach internetowych: Wykorzystanie danych sekwencji trajektorii kursora i przybliżonych obszarów zainteresowania

Coraz większe zastosowanie ankiet internetowych w badaniach nauk społecznych przyniosło ze sobą wyzwanie skutecznego identyfikowania i zarządzania nieuważną/nieostrożnym odpowiadaniem. Istniejące metody wykrywania wykazały ograniczone możliwości, co wskazuje na konieczność wypracowania lepszych metodologii. Niniejsze badanie przedstawia nowatorskie podejście, które wykorzystuje dane sekwencji czasowych akcji związanych z ruchem kursora myszy i stosuje modele głębokiego uczenia maszynowego w celu wykrywania nieuważnego odpowiadania. Wprowadzono koncepcję przybliżonych obszarów zainteresowania (Approximate Areas of Interest - AAOI), wraz z zastosowaniem modeli GRU (Gated Recurrent Units) i BiLSTM (Bidirectional Long Short-Term Memory). Niniejsze badanie prezentuje elastyczne i efektywne narzędzie, które może być stosowane w różnych skalach i kontekstach ankietowych. Wyniki demonstrowane przez proponowane podejście wykazują wyższą dokładność w identyfikacji nieuważnego odpowiadania w ankietach, osiągając do 95% dokładności (*accuracy*) w przypadku testowania na danych eksperymentalnych z wywołaną nieuważnością. Prezentowane podejście stanowi potencjalny przełom w pokonywaniu wszechobecnego wyzwania wykrywania nieostrożnego odpowiadania w ankietach komputerowych.

TOMASZ ŻÓŁTAK, INSTYTUT FILOZOFII I SOCJOLOGII PAN

Style odpowiedzi na pytania samoopisowe jako źródło obciążenia wyników regresji (*latentnej*)

W wystąpieniu podjęty zostanie problem zagrożeń dla trafności analiz regresji (*latentnej*), jakie potencjalnie może stanowić występowanie zróżnicowanych stylów odpowiedzi na pytania samoopisowe w sondażach. Jeśli skłonność do korzystania ze skali odpowiedzi w określony sposób jest skorelowana z mierzoną cechą oraz z interesującą badacza zmienną zależną, której związek z mierzoną cechą chce on badać, pominięcie w modelowaniu stylów odpowiedzi będzie wprowadzać do wyników analiz obciążenie. Problem ten zostanie zaprezentowany w odniesieniu do dwóch modeli psychometrycznych wykorzystywanych najczęściej do uwzględnienia zróżnicowania stylów odpowiedzi wśród respondentów: tzw. IRTrees oraz MNRM (Multidimensional Nominal Response Models). W wystąpieniu przedstawione zostaną wyniki badania symulacyjnego uwzględniającego kwestie możliwości wykrywania modelu generującego dane oraz konsekwencji zastosowania do analizy modelu niezgodnego z modelem generującym - w szczególności modelu całkowicie pomijającego zróżnicowanie stylów odpowiedzi. Uzupełnienie badania symulacyjnego stanowić będzie prezentacja wyników analiz z wykorzystaniem danych z międzynarodowego badania umiejętności osób dorosłych PIAAC.

Problematyka sprawdzianów uwagi (*attention checks*) w ankietach – czego musimy się dowiedzieć, by używać ich skutecznie?

Wystąpienie ma na celu zaprezentowanie tematyki *sprawdzianów uwagi* - specyficznych pytań i zadań, które są stosowane w ankietach (i innych pomiarach) w celu zidentyfikowania nieuważnych uczestników.

Przykładami *testów uwagi* stosowanych w praktyce badawczej są pozycje kwestionariusza typu: „Jeśli przeczytałeś uważnie, zaznacz w tym pytaniu «Zdecydowanie się zgadzam»”, „Pomarańcza to owoc” lub „Byłem na Księżycu”. Oczywiście dla tych pozycji logicznie poprawny jest tylko jeden typ odpowiedzi – jeśli respondent jej nie udzieli, np. zaznaczy jakąkolwiek inną odpowiedź niż „Zdecydowanie się zgadzam”, nie zgadza się, że pomarańcza jest owocem lub twierdzi, że odwiedził księżyc, jest to wyraźny i obiektywny sygnał odpowiadania nieuważnego.

Rzeczywistość nie jest jednak tak różowa. Odsetek respondentów, którzy nie zdali testu uwagi, może być zatrważająco wysoki (nawet do 50, 60, a nawet 87%!) Co więcej, nawet połowa uczestników, którzy nie zaliczą *sprawdzianu uwagi*, może to zrobić z innych powodów niż nieuważne odpowiadanie, np. mylna interpretacja sprawdzianu lub niechęć do poddania się takiemu sprawdzianowi. Istnieją również dowody na to, że niektórzy uczestnicy mogą niekorzystnie reagować na sprawdziany uwagi, np. celowo zaznaczać w nich absurdalne odpowiedzi. Nie ma również konsensusu co do tego, że wykluczenie uczestników, którzy nie zdali *kontroli uwagi*, może znacząco poprawić jakość danych pod jakimkolwiek względem. Dowody te stawiają pewne zastrzeżenia przed stosowaniem kontroli uwagi w celu odfiltrowania nieuważnych uczestników w ankietach internetowych.

Jednak *kontrole uwagi* mają pewne niezaprzeczalne i unikalne zalety: są szybkie i łatwe w użyciu i wciąż łatwiejsze w interpretacji niż inne metody proponowane w celu wyeliminowania nieuważnych uczestników. Ostatnie badania wskazują, że przy odpowiednim zastosowaniu niektóre rodzaje sprawdzianów uwagi mogą okazać się przydatne. Wydaje się, że problemy z kontrolami uwagi wynikają z niedostatecznego zrozumienia, w jaki sposób respondenci je przetwarzają, interpretują i udzielają na nie odpowiedzi. Ponadto zaproponowano wiele różnych rodzajów kontroli uwagi, ale tylko niektóre były szeroko stosowane i testowane.

Wystąpienie ma na celu zaprezentowanie najnowszych odkryć w obszarze stosowania sprawdzianów uwagi w ankietach, a także przedyskutowanie dalszych badań, które są potrzebne do zrozumienia tego, co naprawdę dzieje się, gdy konfrontujemy badanych ze sprawdzianami uwagi. Chciałbym się również zastanowić nad nowymi typami sprawdzianów uwagi, które oferowałyby jak najtrafniejszą identyfikację nieuważnych respondentów.

KATARZYNA CHYL, INSTYTUT BADAŃ EDUKACYJNYCH

ARTUR POKROPEK, INSTYTUT BADAŃ EDUKACYJNYCH

Konstrukcja Testu Czytania dla Dorosłych: wyzwania i problemy

W mojej prezentacji opowiem o badaniu pilotażowym Testu Czytania dla Dorosłych, przeprowadzonym w trzech etapach między kwietniem a lipcem 2023. Test Czytania dla Dorosłych (nazwa robocza) to narzędzie w skład którego wchodzi 1) Testy Rozumienia, 2) Testy Dekodowania, oraz 3) bateria pytań ankietowych wybranych z Kwestionariusza Historii Czytania dla Dorosłych (ARHQ; Bogdanowicz i in., 2015) oraz Kwestionariusza PIAAC (OECD, 2012). Celem badania jest opracowanie testu komputerowego, który trafnie i rzetelnie oceni umiejętność funkcjonalnego czytania osób o niskich kompetencjach (PIAAC: poziom 1 i poniżej poziomu 1).

W etapie I przeprowadziłam spotkania *on-site*, na których osoby badane rozwiązywały test w formie wywiadów poznawczych (*cognitive interviews*). Do spotkania zaproszono osoby z wykształceniem maksymalnie *średnim*, niestudiujące, z grup wiekowych 18-25 (N = 7) oraz 55-62 lat (N = 10). Umożliwiło to ocenę jakościową testu i wyeliminowanie podstawowych błędów przed etapem ilościowym. W etapie II przeprowadziłam spotkania online, na których osoby badane samodzielnie, choć pod nadzorem i z udostępnionym ekranem, rozwiązywały jedną z siedmiu równoległych wersji testu (*Youden square design*). W tych spotkaniach wzięły udział osoby w wieku 19-61 lat (N = 29) z różnym wykształceniem. W etapie III rozesłano linki do osób zrekrutowanych ze względu na zróżnicowany wiek, wykształcenie i miejsce zamieszkania, które test wypełniały samodzielnie i bez nadzoru (N = 102).

Celem analizy danych ilościowych po pilotażu (etap II i III, N = 131 osób) jest opracowanie jednej wersji testu, w skład której wejdą najlepsze teksty i pytania do tekstów, a także najmocniej powiązane z nimi pytania kwestionariuszowe (pośredni pomiar trafności testu) i testy dekodowania. W tym celu wykorzystamy Item Response Theory (IRT). Poprzez zastosowanie IRT, będziemy w stanie oszacować poziom trudności każdego pytania w teście oraz zdolności badanych na podstawie ich odpowiedzi. Dzięki temu będziemy mieli możliwość wybrania najlepszych zadań, różnicujących osoby o niskich kompetencjach czytelniczych. W mojej prezentacji opowiem o strategii analitycznej i wynikach analizy danych którą przeprowadzimy w sierpniu i wrześniu. Zaprezentuję narzędzie, i, licząc na głosy w dyskusji, przedstawię bieżące trudności wynikłe przed rozpoczęciem badania głównego.

SESJA 3. Prowadząca: Katarzyna M. Wyrzykowska (IFiS PAN)

PIOTR JABKOWSKI, WYDZIAŁ SOCJOLOGII, UAM W POZNANIU

ANETA PIEKUT, SHEFFIELD METHODS INSTITUTE, SHEFFIELD UNIVERSITY

Nielosowy i nieignorowalny. Mechanizm błędu uchylania się od odpowiedzi na pytanie o dochód w Europejskim Sondażu Społecznym, 2008-2018

Prezentacja poświęcona jest konsekwencjom uchylania się respondentów od odpowiedzi na pytanie o całkowity dochód gospodarstwa domowego w badaniach Europejskiego Sondażu Społecznego. Rozpoczniemy od prezentacji modelu teoretycznego, w którym rozpoznamy dwa mechanizmy odpowiedzialne za uchylanie się od odpowiedzi na pytanie o dochód a mianowicie: a) złożoność zadania jakim jest konieczność kalkulacji całkowitego dochodu wszystkich członków gospodarstwa domowego oraz b) drażliwość pytania będąca konsekwencją wypytywania się przez ankietera o dochody wszystkich osób zamieszkujących z respondentem w jednym gospodarstwie domowym. W analizie wykorzystamy wyniki pochodzące z 18-tu krajów uczestniczących w latach 2008-2018 we wszystkich kolejnych sześciu rundach Europejskiego Sondażu Społecznego. Uzyskane wyniki modelowania wielopoziomowego (z respondentami, ankietarami i krajami konstytuującymi różne poziomy analizy) pokażą, że złożoność zadania prowadzi do częstszego udzielania odpowiedzi „Nie wiem”, podczas gdy drażliwość pytania do częstszego odmawiania udzielenia odpowiedzi na zadane respondentom pytanie. Pokażemy ponadto, że skłonność do odmowy maleje wraz z wielkością dochodu, podczas gdy skłonność do udzielania odpowiedzi „Nie wiem” wzrasta. Z tego ostatniego wywnioskujemy, że mechanizm stojący za uchylaniem się od odpowiedzi na pytanie o dochód jest powiązany z poziomem dochodu i ma charakter nielosowy. Pokażemy również, że nieudzielenie odpowiedzi na pytanie o dochód ma poważne konsekwencje nie tylko dla jakości pomiaru samego dochodu, ale również dla jakości pomiaru innych zmiennych zawartych w kwestionariuszu ESS.

PAWEŁ PENSZKO, INSTYTUT BADAŃ EDUKACYJNYCH

Zapewnianie reprezentatywności próby - doświadczenia z realizacji badania PIAAC

PIAAC jest międzynarodowym badaniem umiejętności osób dorosłych w zakresie rozumienia tekstu, rozumowania matematycznego i rozwiązywania problemów wymagających adaptacji do zmieniającej się sytuacji. Stosowanych jest w nim wiele standardów i procedur, służących uzyskaniu dobrej jakości, reprezentatywnych i porównywalnych między krajami danych wynikowych. Należy do nich na przykład ustalenie minimalnego poziomu efektywności realizacji próby (wskaźnika *response rate*, RR) na 50%. Próg ten udało się przekroczyć w Polsce, gdy w kwietniu 2023 roku zakończyła się realizacja wywiadów i pomiarów umiejętności w ramach drugiego cyklu badania PIAAC. Z zebranych w jej trakcie danych wynika szereg wniosków istotnych w kontekście zapewnienia reprezentatywności badań kwestionariuszowych w naukach społecznych. Dane te pozwalają oszacować, że co najmniej 18% adresów widniejących w rejestrze PESEL jest nieaktualnych z powodu przeprowadzki mieszkańca, w tym w przypadku co najmniej 6% - w ramach Polski. W badaniu zastosowano wiele środków, które miały sprzyjać osiągnięciu wysokiego RR. Spośród nich bardzo skuteczne, ale zarazem bardzo kosztowne finansowo okazało się uzależnienie wysokości wynagrodzenia wykonawcy prac terenowych od uzyskanego RR.

Nieduże liczebnie, ale wymierne rezultaty przyniosły próby pozyskania aktualnego adresu respondenta i konwersja odmów. System priorytetyzacji adresów wydawał się mieć pewien wpływ na zachowanie ankietów, jego oddziaływanie na RR jest skomplikowane. Brak danych pozwalających porównać skuteczność stosowanych zachęt dla respondentów, takich jak bon zakupowy czy możliwość udziału w konkursie, dają się natomiast zauważyć różnice między nimi pod względem odsetków respondentów, którzy pamiętają o propozycji otrzymania zachęty danego rodzaju i którzy skorzystali z niej. Doświadczenia z drugiego cyklu badania PIAAC pokazują, że stosowanie techniki CASI, przy której respondent sam udziela odpowiedzi za pomocą tabletu, nie musi oznaczać w praktyce wyłączenia z badania istotnej części populacji mieszkańców Polski. Rzadko stosowanym w badaniach kwestionariuszowym rozwiązaniem jest ankieta obserwacyjna, w której ankietę odnotowuje informacje o budynku znajdującym się pod wskazanym mu adresem. Analiza danych z takiej ankiety, przeprowadzona z zastosowaniem kryteriów kompletności, trafności, zgodności między obserwatorami i wartości predykcyjnej, wskazuje na to, że mogą być one przydatne przy weryfikacji reprezentatywności zrealizowanej próby lub tworzeniu wag korekcyjnych. Pewnym wyzwaniem jest natomiast zapewnienie wysokiej jakości danych z ankiety obserwacyjnej, w czym może pomóc wyszkolenie ankietów i przekazanie im precyzyjnych kryteriów klasyfikacji.

KRZYSZTOF BULKOWSKI, INSTYTUT BADAŃ EDUKACYJNYCH / WYDZIAŁ SOCJOLOGII UW

Innowacje metodologiczne w międzynarodowych badaniach edukacyjnych.

Przypadek piątej edycji badania PIRLS

Międzynarodowe porównawcze badania edukacyjne pozwalają na ocenę umiejętności uczniów, uwzględniając zarazem szerszy kontekst uwarunkowań indywidualnych, szkolnych i związanych z systemem edukacji. Ich projektowanie, organizacja i realizacja jest dużym wyzwaniem, w które zaangażowanych jest wiele osób. Na przestrzeni ponad 50 lat realizacji takich badań udoskonalane są procedury pomiarowe zapewniające porównywalność wyników. Obecnie badania te charakteryzują się też wysokim stopniem szczegółowości planowania całego procesu badawczego na każdym jego etapie. Wystąpienie dotyczy będzie Międzynarodowego Badania Postępów Biegłości w Czytaniu – PIRLS 2021 (*Progress in International Reading Literacy Study*), którego wyniki zostały opublikowane na całym świecie w maju 2023 r.

Piąta edycja PIRLS, przeprowadzana 20 lat po rozpoczęciu badania, charakteryzowała się m.in. wdrożeniem dwóch innowacji metodologicznych. Pierwszy raz w badaniach prowadzonych przez Międzynarodowe Stowarzyszenie Mierzenia Osiągnięć Szkolnych (IEA) zastosowano quasi-adaptatywny schemat przydziału zadań (Group Adaptive Design). Ze względu na stwierdzone w poprzednich edycjach PIRLS duże zróżnicowanie umiejętności czytelnicych uczniów pomiędzy krajami oraz fakt, że przy losowym, acz równomiernym przydziale zadań w każdym z krajów otrzymywano mniej dokładne oszacowania umiejętności uczniów, postanowiono wdrożyć schemat, który poprawiłby precyzję oszacowań. Ze względu na realizację badania w wielu krajach w formie papierowej nie można było wdrożyć pełnej adaptatywności (czyli wyświetlania się zadań o określonej trudności na podstawie wcześniejszych odpowiedzi ucznia). Dlatego zastosowano schemat quasi-adaptatywny, którego parametry ustalane były na poziomie kraju. Każdemu z krajów, na podstawie wyników wcześniejszej edycji badania PIRLS, przypisano odpowiedni schemat przydziału zeszytów testowych. W krajach o wyższych wynikach odsetek przydzielonych uczniom zeszytów testowych określanych jako ""trudne"" był wyższy.

Drugą innowacją zastosowaną w PIRLS było, wdrożone już wcześniej w innych badaniach IEA (np. w badaniu TIMSS 2019), przeniesienie narzędzi badawczych, czyli tekstów, zadań testowych oraz pytań kwestionariuszowych, z wersji papierowych do wersji wyświetlanych na urządzeniach cyfrowych. Na taki sposób prowadzenia badania wśród uczniów w czwartym roku nauki szkolnej zdecydowało się 26 z 57 biorących udział w badaniu krajów. Zmiana wiązała się z koniecznością przeprowadzenia dodatkowego badania za pomocą narzędzi papierowych (*Bridge Study*) w celu opracowania statystycznego „modelu przejścia”, dzięki któremu zapewniona będzie porównywalność wyników pomiędzy wersją papierową i wersją elektroniczną.

Celem wystąpienia jest omówienie założeń zastosowanych innowacji i przedstawienie ich wyników. O ile wyniki pierwszej z nich były zgodne z oczekiwaniami jej projektantów – zastosowanie schematu quasi-adaptatywnego zwiększyło precyzję pomiaru – o tyle druga z innowacji, polegająca na zastosowaniu narzędzi w wersji elektronicznej, spowodowała konieczność zmiany podejścia analitycznego oraz przesunięcie terminu publikacji wyników badania o pół roku.

W trakcie wystąpienia przedstawiony zostanie również ważny kontekst, który posłużył do uzasadnienia opóźnień w publikacji wyników – badanie PIRLS było jedynym międzynarodowym badaniem edukacyjnym, które zdecydowano się przeprowadzić w szkołach w trakcie trwania pandemii. Decyzje organizatorów badania wpłynęły nie tylko na jego wyniki, ale także na brak możliwości spełnienia głównego celu badania (uzyskanie porównywalności z innymi krajami i z wcześniejszymi pomiarami) w 14 z 57 biorących udział w badaniu krajów. W wystąpieniu dokonana zostanie próba zdiagnozowania przyczyn powstałych problemów poparta sformułowaniem rekomendacji mających na celu uniknięcie podobnych problemów w przyszłości.

NATALIA NADOLNA, UNIWERSYTET KARDYNAŁA STEFANA WYSZYŃSKIEGO W WARSZAWIE

ZUZANNA SZALEK, UNIWERSYTET KARDYNAŁA STEFANA WYSZYŃSKIEGO W WARSZAWIE

MARTYNA ANDRYSZKIEWICZ, UNIWERSYTET ŁÓDZKI

MACIEJ GRZEGORCZYK, UNIWERSYTET KARDYNAŁA STEFANA WYSZYŃSKIEGO W WARSZAWIE

MARIA JASTRZĘBSKA, UNIWERSYTET KARDYNAŁA STEFANA WYSZYŃSKIEGO W WARSZAWIE

MAGDALENA PUCHALSKA, AKADEMIA PEDAGOGIKI SPECJALNEJ IM. MARII GRZEGORZEWSKIEJ W WARSZAWIE

PRZEMYSŁAW SZABŁOWSKI

JOANNA ŚWIDERSKA, AKADEMIA PEDAGOGIKI SPECJALNEJ IM. MARII GRZEGORZEWSKIEJ W WARSZAWIE

Próby badawcze w badaniach psychologicznych – oczekiwania versus rzeczywistość

Co roku ukazuje się wiele polskich publikacji z zakresu psychologii zarówno w krajowych, jak i zagranicznych czasopismach. Ważnym elementem każdego badania empirycznego, pozwalającym na ocenę jakości uzyskanych rezultatów, jest opis próby badawczej. Dlatego zadaliśmy sobie pytanie, co powinno zawierać się w tych opisach oraz czy publikowane badania zawierają wystarczającą ilość informacji, aby można było dokonać oceny metodologicznej procedury

badawczej (m.in. w kontekście replikacji badań). W tym celu dokonaliśmy przeglądu badań empirycznych opublikowanych przez polskich autorów w ostatnich dwóch latach (2021-2022) w czterech czasopismach: Personality and Individual Differences, Frontiers in Psychology, Roczniki Psychologiczne i Polskie Forum Psychologiczne. Analizując N = 297 opisów prób zwracaliśmy uwagę na takie kryteria, jak: metoda doboru próby, wielkość próby, uzasadnienie wielkości próby, docelowa populacja, kryteria włączenia i wyłączenia osób badanych, odsetek osób wykluczonych, forma badania (stacjonarne, online, mieszane), procedura badania, kiedy badanie zostało przeprowadzone, kto prowadził badanie, wynagrodzenie dla badanych oraz elementy charakterystyki respondentów. Celem wystąpienia jest zaprezentowanie wyników i wniosków z przeprowadzonych analiz oraz zaproponowanie rekomendacji w zakresie formułowania wyczerpującego opisu próby, a także podzielenie się doświadczeniami zdobytymi na bazie tego przeglądu.

GRZEGORZ HUMENNY, INSTYTUT BADAŃ EDUKACYJNYCH

MAŁGORZATA KŁOBUSZEWSKA, INSTYTUT BADAŃ EDUKACYJNYCH

Rola dyplomu zawodowego w początkach kariery zawodowej absolwentów techników. Analiza z wykorzystaniem modelowania wielopoziomowego z uwzględnieniem problemu selekcji do próby. Dowody na podstawie danych administracyjnych

W tym referacie poruszamy kwestię roli dyplomu zawodowego, certyfikatów kwalifikacji zawodowych oraz doświadczenia zawodowego zdobytego w czasie edukacji dla sytuacji na rynku pracy absolwentów techników w Polsce. Uwzględniamy także charakterystyki zawodów, w których kształcili się absolwenci oraz uwarunkowania lokalnych rynków pracy. Wykorzystując teorie ekonomiczne dotyczące kapitału ludzkiego i *signalingu*, badamy, czy te aspekty edukacji zawodowej wpływają na prawdopodobieństwo zatrudnienia i poziom zarobków absolwentów. W tym celu wykorzystujemy polskie dane rejestrowe pochodzące z Systemu monitorowania karier absolwentów szkół ponadpodstawowych. Ze względu na charakter naszych danych i problematykę badawczą stosujemy modelowanie wielopoziomowe uwzględniające jednocześnie krzyżowe zagnieżdżenie absolwentów w zawodach, których się uczyli oraz w powiatach, na terenie których zlokalizowana była szkoła do której uczęszczali (typ *crossclassified* w Mplus), a problem selekcji do próby rozwiązujemy wykorzystując *korektę Heckmana*.

Nasze wyniki wskazują, że uzyskanie dyplomu zawodowego lub nawet części certyfikatów kwalifikacji zawodowych wiąże się z większym prawdopodobieństwem zatrudnienia i wyższymi zarobkami. Zdobycie doświadczenia zawodowego w trakcie nauki w szkole średniej ma jeszcze większe znaczenie dla znalezienia pracy, ale jego znaczenie może zmniejszać się wraz z upływem czasu, podczas gdy efekt dyplomu zawodowego jest bardziej stabilny. Znaczenie mają również charakterystyki nauczanych zawodów (np. stopień feminizacji) oraz sytuacja na lokalnym rynku pracy. Nasze wyniki mogą być cennymi informacjami dla polityki edukacyjnej, ponieważ oceniają skuteczność wprowadzonych przez poprzednie reformy edukacyjne zmian, tj. formalnego potwierdzenia kwalifikacji zawodowych w systemie szkolnym.

22 września

SESJA 4. Prowadzący: Henryk Domański (IFiS PAN)

JOLANTA PEREK-BIAŁAS, UNIwersYTET JagIELLOŃSKI/SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE

PAULINA SKÓRSKA, UNIwersYTET JagIELLOŃSKI

MILENA MAJ, UNIwersYTET JagIELLOŃSKI

Właściwości skali AFCCQ – walidacja w grupie osób starszych w Polsce

Celem wystąpienia będzie zaprezentowanie walidacji narzędzia do badania oceny przez osoby starsze w Polsce aspektów powiązanych z jakością życia zgodnie z wytycznymi Światowej Organizacji Zdrowia. Oryginalny kwestionariusz *Miasta i Społeczności Przyjazne Bez Względu na Wiek (AFCCQ)* (Dikken et al., 2020) opracowano w wersji polskiej i walidowano w ramach międzynarodowego, interdyscyplinarnego projektu badawczego pt. „Miasto i Współpraca: Starsze osoby współtworzące zrównoważone i przyjazne bez względu na wiek miasto” (ang. „*City&Co: Older Adults Co-Creating a Sustainable Age-friendly City*”) - Polska, Holandia, Rumunia).

Kwestionariusz bada różne aspekty jakości życia w mieście, m.in. takie jak: dostępność infrastruktury dla osób starszych i z niepełnosprawnościami, dostępność informacji, uczestnictwo w życiu towarzyskim i społecznym, dostęp do usług medycznych, sytuację finansową osób starszych.

W toku prezentacji przedstawimy kształt, zastosowanie i właściwości psychometryczne narzędzia. W badaniu w Polsce w 2023 roku wzięło udział 801 mieszkańców Krakowa i Wrocławia, w wieku powyżej 65 lat. Badanie było realizowane w oparciu o próbę losową, odzwierciedlającą strukturę według płci i wieku dzielnic obu miast.

W czasie prezentacji zaprezentowane zostaną wyniki dotyczące trafności teoretycznej modelu, na którym opiera się konstrukcja skali. Odniesiemy się do rezultatów confirmacyjnej analizy czynnikowej (CFA) oraz do dopasowania modelu do danych (na podstawie wskaźników takich jak: RMSEA, CFI, TLI, czy SRMR). Podsumujemy rzetelność skali w oparciu o mierniki alfa Cronbacha, oraz omega McDonalda. Trafność zbieżna została w badaniu oceniona na podstawie metod przeciętnej wyodrębnionej wariancji (*Average Variance Extracted, AVE*) oraz rzetelności kompozytowej (*Composite Reliability, CR*). Do analizy trafności dyskryminacyjnej została użyta macierz HTMT (*Heterotrait-monotrait ratio of the correlations, HTMT*).

Na koniec zostaną przedstawione wyzwania związane z przygotowaniem skali w polskich warunkach, w tym wskazanie stabilności funkcjonowania skali.

Literatura:

Dikken, J., van den Hoven, R. F., van Staalduinen, W. H., Hulsebosch-Janssen, L. M., & Van Hoof, J. (2020). How older people experience the age-friendliness of their city: Development of the age-friendly cities and communities questionnaire. *International journal of environmental research and public health*, 17(18), 6867.

BARBARA KOWALCZYK, KOLEGIUM ANALIZ EKONOMICZNYCH, SZKOŁA GŁÓWNA
HANDLOWA W WARSZAWIE

EMILIA TOMCZYK, KOLEGIUM ANALIZ EKONOMICZNYCH, SZKOŁA GŁÓWNA HANDLOWA W
WARSZAWIE

Porównanie metod pomiaru konsensusu w badaniach koniunktury

W referacie porównujemy alternatywne metody oceny konsensusu w badaniach koniunktury, w których respondenci wyrażają oczekiwania na skali porządkowej. Wiarygodna metoda pomiaru siły konsensusu w oczekiwaniach respondentów dostarczyłaby badaczom cennych informacji, stanowiąc wiodący wskaźnik nastrojów podmiotów gospodarczych. Nie ma jednak jednej ogólnie przyjętej miary matematycznej służącej do oceny zgodności między wyrażanymi przez respondentów opiniami. W literaturze spotkać można kilka podejść do tego problemu, opartych m.in. na statystycznych miarach dyspersji, entropii Shannona czy na pojęciu k -wymiarowego sympleksu. W referacie przedstawiamy zdefiniowane w literaturze miary konsensusu oraz omawiamy ich zalety i ograniczenia. Następnie wykorzystujemy te wskaźniki do analizy oczekiwań respondentów wyrażonych w badaniach koniunktury w przetwórstwie przemysłowym w Polsce i porównujemy wyniki dla różnych zmiennych ekonomicznych. W kilku przypadkach znajdujemy powtarzalne schematy w zachowaniu miar konsensusu: oczekiwania cenowe charakteryzują się najwyższym stopniem konsensusu, a oczekiwania na temat produkcji i zamówień – najniższym. Wskazujemy również powiązania między stopniem konsensusu a stopniem optymizmu wśród respondentów mierzonym statystykami bilansowymi w przypadku cen, zatrudnienia i ogólnej sytuacji gospodarczej.

Słowa kluczowe: badania koniunktury, oczekiwania, dane jakościowe, konsensus, przetwórstwo przemysłowe

PAWEŁ MATUSZEWSKI, COLLEGIUM CIVITAS

Wyszukiwanie złożonych zjawisk społecznych w dużych i nieustrukturyzowanych danych tekstowych

Prezentacja ma na celu opisanie wspomaganą komputerowo metody wyszukiwania złożonych zjawisk społecznych w nieustrukturyzowanych danych tekstowych. Zbieranie takich danych z Internetu zwykle polega na używaniu słów kluczowych do znalezienia istotnych dokumentów. Choć ten krok ma kluczowe znaczenie dla uzyskania wyników, w licznych badaniach badacze wybierają słowa kluczowe arbitralnie, nie podejmując prób ewaluacji tego, na ile zastosowany zbiór prowadzi do poprawnej selekcji. Skala możliwych błędnych klasyfikacji – zarówno fałszywie pozytywnych jak i fałszywie negatywnych - które w ten sposób można popełnić wzrasta wraz z złożonością poszukiwanego konceptu.

Opracowana przez mój zespół procedura zostanie zademonstrowana na przykładzie teorii spiskowych a więc relatywnie skomplikowanego zjawiska. Teorie spiskowe nie mają sztywnych granic koncepcyjnych, ewoluują, a osoby rozpowszechniające teorie spiskowe często używają specyficznego języka, aby uniknąć cenzury lub algorytmów platform, które je blokują. Z tego względu teksty, które je zawierają są szczególnie trudne do wychwycenia.

Procedura opiera się na połączeniu word embedding z ludzkimi zdolnościami poznawczymi. Opracowany indeks radził sobie w wykrywaniu teorii spiskowych w dużym korpusie tekstów (~1,1 miliona tweetów) lepiej niż nadzorowane algorytmy uczenia maszynowego (*regresja lasso*,

XGBoost, SVM i *konwolucyjna sieć neuronowa*) a także lepiej w porównaniu do zbioru słów kluczy, które zostały arbitralnie wybrane przez grupę uczestników badania ankietowego (N = 154).

KAZIMIERZ CZARNOCKI

Zastosowanie interpretowalnych modeli uczenia maszynowego do lepszej i prostszej specyfikacji modeli parametrycznych

Cel:

Celem niniejszej prezentacji jest pokazanie, jak interpretowalne modele uczenia maszynowego (UM)¹ mogą zostać wykorzystane do lepszej (tj. lepiej pasującej do danych) i prostszej (tj. mniej czasochłonnej i mniej podatnej na błędy) specyfikacji modeli parametrycznych.

Istotność:

Poprawna specyfikacja modelu jest kluczowa dla jakości analizy oraz wiarygodności wniosków. Przykładowo Vowels (2021) argumentuje, że niepoprawna specyfikacja modeli jest istotnym czynnikiem leżącym u podstaw kryzysu replikacji w naukach społecznych. Dzieje się tak, ponieważ rezultatem niepoprawnej specyfikacji modelu są obciążone oszacowania parametrów. Ponadto, opisana procedura jest mniej pracochłonna od manualnej specyfikacji modelu parametrycznego. Więc, może oszczędzić czas badaczy.

Abstrakt:

Zanim model parametryczny zostanie użyty, żeby odpowiedzieć na pytanie badawcze, musi zostać odpowiednio wyspecyfikowany. Szczegóły specyfikacji zależą od danego modelu, ale w uogólnieniu proces specyfikacji może być przedstawiony jako wybór zbioru operacji na danych wejściowych (ang. *input*) modelu, który maksymalizuje dopasowanie modelu do danych (np. efekty kwadratowe, interakcje, log transformacje).²

Niestety specyfikacja modelu parametrycznego jest pracochłonnym i podatnym na błędy procesem. Pomimo, że istnieją testy specyfikacji, to nie znajdują one optymalnej specyfikacji. W konsekwencji, nie jest jasne jak można przejść od początkowej specyfikacji do optymalnej. W praktyce badacze stosują iteratywną procedurę, która jest pracochłonna i nie gwarantuje optymalnych wyników.

Modele UM nie wymagają manualnej specyfikacji. Automatycznie uwzględniają interakcje i nieliniowości. Innymi słowy, bezpośrednio przechodzą od danych do finalnej specyfikacji. Ich wadą jest to, że mechanizm predykcyjny nie jest łatwy do zrozumienia. W konsekwencji, nawet jeśli model poprawnie wykrył wszystkie zależności w danych, to jest bardzo trudno wydobyć z niego mechanizm, który mógłby zostać użyty do specyfikacji modelu parametrycznego.³

Techniki interpretowalnego UM mają na celu stworzenie wyjaśnienia, jak dany model UM stworzył predykcję. Ich wynik (ang. *output*) może zostać użyty do znalezienia interakcji i nieliniowości. Z powyższego, jeśli założymy, że model UM dobrze pasuje do danych, to może zostać on użyty do specyfikacji modelu parametrycznego.

Niniejsza prezentacja zademonstruje przydatność interpretowalnych modeli UM w naukach społecznych na przykładzie *lasów losowych* (Breiman, 2001) jako modelu UM, oraz wartości Shapley'a (Merrick & Taly, 2020) jako techniki interpretacji. Ta kombinacja tworzy ramę

¹ Interpretowalny model UM to taki model, którego wynik może zostać zrozumiany i zinterpretowany przez człowieka.

² To uproszczenie. Przykładowo, nadmierne dopasowanie (ang. *overfit*) jest również ważną kwestią.

³ Ten paragraf znacząco upraszcza i generalizuje. Problematykę rozwijają Bishop & Nasrabadi (2006) oraz Murphy (2012).

teoretyczną, która przechodzi bezpośrednio od danych do modelu i może być łatwo przeanalizowana w celu znalezienia poprawnej specyfikacji modelu parametrycznego.

Literatura:

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Merrick, L., & Taly, A. (2020). The explanation game: Explaining machine learning models using shapley values. In *Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4* (pp. 17-38). Springer International Publishing.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Vowels, M. J. (2021). Misspecification and unreliable interpretations in psychology and social science. *Psychological Methods*.

MAGDALENA KOCHAŃSKA, AKADEMIA PEDAGOGIKI SPECJALNEJ IM. MARII GRZEGORZEWSKIEJ W WARSZAWIE

NATALIA BANASIK-JEMIELNIAK, AKADEMIA PEDAGOGIKI SPECJALNEJ IM. MARII GRZEGORZEWSKIEJ W WARSZAWIE

Badania psychologiczne dzieci prowadzone online – korzyści i wyzwania

Prowadzenie badań przez Internet w naukach społecznych, w tym psychologicznych, nie jest niczym nowym. Medium to jest wygodnym pośrednikiem zarówno w wypadku badań kwestionariuszowych, jak i eksperymentalnych, ilościowych i jakościowych (Batorski, Olcoń-Kubicka, 2006). Jednak w badaniach naukowych z wykorzystaniem testów w czasie rzeczywistym czy w diagnozie psychologicznej dziecka do czasu pandemii Covid-19 dominowała praktyka badania w formie “on-site” (Tucker, 2012). Czas epidemii koronawirusa obnażył pilną potrzebę adaptacji testów psychologicznych i narzędzi badawczych do warunków testowania online (Nelson, Plante, 2022). I choć instytucje takie jak *American Psychological Association* (2020) zalecały początkowo powstrzymanie się przed używaniem narzędzi, które nie zostały zatwierdzone do użytku w formie online, to coraz więcej badań pokazuje dużą wiarygodność badań psychologicznych w wersji komputerowej za pośrednictwem Internetu (Peña, Sutherland, 2022).

W prowadzonym na Akademii Pedagogiki Specjalnej projekcie metodologicznym dotyczącym dzieci dwujęzycznych (polsko-angielskich i polsko-niemieckich), sprawdzamy różne warunki badania zadaniem powtarzania zdań (SRep, ang. Sentence Repetition Task) (Przygocka i in., 2021), w tym porównujemy rzetelność badań prowadzonych online i on-site. SRep jest narzędziem, które pozwala na rzetelną ocenę umiejętności morfoskładniowych (gramatycznych) dziecka (Polišenská et al., 2015). Jest to zadanie wymagające rozumienia i produkcji języka, a zatem można założyć, że potencjalne problemy wynikające ze specyfiki badania online mogą znacznie wpływać na wyniki dzieci (Manning i in., 2020).

Podczas wystąpienia przedstawione zostaną wnioski, jakie wyciągnęliśmy z prowadzonych od roku badań dzieci za pośrednictwem Internetu. Omówione zostaną korzyści i wyzwania, jakie wiążą się z tą formą testowania m.in. w aspektach: etycznym, standaryzacyjnym, technologicznym i ekonomicznym. Wnioski te wynikają nie tylko z naszych doświadczeń w projekcie, ale także z

analiz pozyskanych danych. Zaprezentowane zostaną analizy wyników 50 dzieci dwujęzycznych i 25 dzieci jednojęzycznych w wieku 5;0–6;11, które były badane tym samym zadaniem SRep w obu warunkach: przez Internet i podczas spotkania *na żywo* z badaczem.

Literatura:

American Psychological Association. (2020). Guidance on psychological tele-assessment during the COVID-19 crisis. [http:// www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19](http://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19)

Batorski, D., Olcoń-Kubicka, M. (2006). Conducting research via the Internet—basic methodological issues. *Sociological Studies*, 2, 106–108.

Manning, B. L., Harpole, A., Harriott, E. M., Postolowicz, K., Norton, E. S. (2020). Taking language samples home: Feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. *Journal of Speech, Language, and Hearing Research*, 63(12), 3982–3990.

Nelson, N. W., Plante, E. (2022). Evaluating the equivalence of telepractice and traditional administration of the Test of Integrated Language and Literacy Skills. *Language, speech, and hearing services in schools*, 53(2), 376–390.

Peña, E. D., Sutherland, R. (2022). Can you see my screen? Virtual assessment in speech and language. *Language, Speech, and Hearing Services in Schools*, 53(2), 329–334.

Polišenská, K., Chiat, S., Roy, P. (2015). Sentence repetition: What does the task measure?. *International Journal of Language & Communication Disorders*, 50(1), 106–118.

Przygocka, S., Banasik-Jemiłniak, N., Smoczyńska, M., Haman, E. (2021). Zadanie Powtarzania Zdąń – polska krótka wersja. Adaptacja LITMUS Sentence Repetition Task. Materiał niepublikowany. Wydział Psychologii, Uniwersytet Warszawski.

Tucker, J. K. (2012). Perspectives of speech-language pathologists on the use of telepractice in schools: Quantitative survey results. *International Journal of Telerehabilitation*, 4(2), 61–72.

SESJA 5. Prowadząca: Kinga Zawadzka (IFiS PAN)

JACEK HAMAN, WYDZIAŁ SOCJOLOGII UW

Podział mandatów między okręgi w wyborach parlamentarnych w Polsce – problemy normatywne i pomiarowe. O wspólnych problemach prawa wyborczego i badań ilościowych

„Reprezentatywność” jest kluczowym pojęciem dla społecznych badań ilościowych; „reprezentatywność” jest również podstawowym wymogiem formułowanym pod adresem instytucji przedstawicielskich. Nie jest to jedynie przypadkowe użycie tego samego słowa – w pewnych przynajmniej aspektach chodzi o dokładnie tę samą „reprezentatywność”, a problemy z jej zapewnieniem wynikają z tych samych źródeł. Dotyczy to w szczególności jakości i charakteru danych demograficznych będących podstawą do określania wielkości okręgów wyborczych (w prawie wyborczym) oraz określania parametrów oraz dobieraniu prób (w badaniach ilościowych). Z drugiej strony, tak, jak w prawie wyborczym istotne jest określenie, jaką społeczność mają reprezentować posłowie, przy przeprowadzaniu badań na próbach reprezentatywnych niezbędna jest refleksja nad tym, jak konkretne decyzje dotyczące procedur doboru próby determinują rzeczywistą populację poddaną badaniu.

Zasady dokonywania podziału mandatów między okręgi do Sejmu, regulowane są przez art. 201-203 i 260-216 Kodeksu Wyborczego oraz wydane na ich podstawie akty wykonawcze. Stanowią one operacjonalizację konstytucyjnej normy równości wyborów do Sejmu. Wspomniane przepisy Kodeksu Wyborczego właściwie zawsze były traktowane jako „techniczne” i nie budzące kontrowersji (poza bardzo wąskim gronem specjalistów), tymczasem wiąże się z nimi szereg problemów normatywnych, prawnych, metodologiczno-pomiarowych oraz politycznych. Te ostatnie nie wchodzą w tematykę naszej konferencji, więc nimi zajmować się na razie nie będę. Kwestie prawne natomiast splatają się z problemami normatywnymi i metodologicznymi (to w formie przepisów prawa operacjonalizowane są normy sprawiedliwości i to przepisy prawa narzucają wybór metod pomiaru zmiennych uwzględnianych przy dzieleniu mandatów).

Określenie zasad metody podziału proporcjonalnego mandatów wymaga rozstrzygnięcia trzech kwestii: wyboru jednej z konkretnych metod operacjonalizujących zasadę proporcjonalności (np. metody Here-Niemayera lub Webstera-Sainte-Laguë), ogólne wskazanie, do czego podział ma być proporcjonalny (np. liczby mieszkańców, liczby uprawnionych do głosowania, liczby obywateli itp.) i wreszcie określenie dokładnej metody wyznaczania (pomiaru) wartości tej zmiennej (moment dokonywania pomiaru, źródło danych – np. dane z rejestrów [jakich?], dane ze spisu powszechnego [z jaką definicją liczby ludności?]). Osobną kwestią jest określenie sposobu egzekwowania prawidłowości (tzn. zgodności z wcześniej wprowadzonymi zasadami) stosowanego w wyborach podziału; to zagadnienie – choć bardzo istotne i interesujące – nie należy jednak do obszaru tematycznego naszej konferencji.

Czy przyjmując konkretne rozwiązania ustawodawca dokonywał świadomego wyboru między alternatywnymi interpretacjami pojęcia proporcjonalności oraz sposobami pomiaru wielkości populacji, czy też przyjmował „pierwsze z brzegu” rozwiązania poszczególnych problemów? Wiele wskazuje, że, niestety, często to drugie. W swoim referacie chcę pokazać, jak można by uzasadnić przyjęcie rozwiązań innych niż obowiązujące obecnie w Polsce i jakie miałyby to konsekwencje dla podziału mandatów między okręgi wyborcze do Sejmu. Ponieważ sama kwestia wyboru metody podziału proporcjonalnego jest już w literaturze dość dokładnie opisana (także przeze mnie), a także luźniej związana jest z tematyką konferencji, skupię głównie się na pozostałych aspektach, a więc wyborze „bazy proporcjonalności” oraz potencjalnych

konsekwencjach różnych podejść do pomiaru jej wielkości. Problemy, które przy tym można napotkać, są analogiczne do problemów, które trzeba rozwiązywać planując badania na próbach reprezentatywnych.

SŁAWOMIR BARTNICKI, INSTYTUT SOCJOLOGII UNIWERSYTET W BIAŁYMSTOKU

MACIEJ ALIMOWSKI, SZKOŁA DOKTORSKA NAUK SPOŁECZNYCH UNIWERSYTETU W BIAŁYMSTOKU

Przewidywanie wyników wyborów organów wykonawczych gmin w nadchodzących wyborach w 2024 roku za pomocą algorytmów uczenia maszynowego i wnioski płynące z modeli black box dla wyjaśniania i rozumienia w naukach społecznych

Funkcja predykcyjna w naukach społecznych jest eksploatowana w stopniu nieporównywalnie mniejszym, niż funkcja wyjaśniająca czy rozumiejąca. W referacie podejmowana jest próba częściowego uzupełnienia tych deficytów poprzez przewidywanie wartości wybranych zdarzeń charakteryzujących nadchodzące wybory do organów wykonawczych gmin w 2024 roku. Przewidywania realizowane będą poprzez zastosowanie wybranych metod uczenia maszynowego począwszy od prymarnych, jak np. naiwne sieci bayesowskie (NB) (Lewis 1998; Domingos i Pazzani 1997; Langley i in. 1992) czy metoda wektorów nośnych (SVM) (Cortes, Vapnik 1995), poprzez lasy losowe (RF) (Ho 1995; Breiman 2001), na modelach uczenia zespołowego H2O (LeDell i in. 2023) kończąc. Do predykcji zostaną wybrane niektóre z rezultatów tych wyborów (zmiennie zależne), spośród wielu możliwych, (np. re-kandydowanie i wygrywanie inkumbentów, liczba kandydatów w nadchodzących wyborach, wystąpienie wyborów plebiscytowych, wystąpienie drugiej tury wyborów, kandydowanie kobiet czy oczekiwana frekwencja wyborcza, itp.). Schemat predykcji obejmuje model treningowy, walidacyjny i testowy w zakresie porównań ze stanem faktycznym w wyborach w 2018 roku, oraz predykcje wyników na nadchodzące wybory. W modelach uwzględnione zostanie ryzyko wpływu ekstraordynaryjnych czynników zewnętrznych zaburzających ustabilizowany horyzont zdarzeń w dotychczasowych wyborach bezpośrednich do 2018 roku. Zakłócenia są skutkiem działań rządu w zakresie: a) ograniczenia dopuszczalnej liczby kadencji na tych stanowiskach; b) wydłużenia czasu trwania kadencji do 5 lat; oraz c) ponownego wydłużenia czasu trwania kadencji w trakcie jej trwania o kolejne 6 miesięcy. Aby uwzględnić wpływ tych zakłóceń na przewidywane wyniki, proponowana jest modyfikacja modeli w zakresie ingerencji w wartości zmiennych oraz prioryzację wynikającą z subiektywizacji popartej wiedzą o dotychczasowej dynamice wyników tych wyborów. Ostatecznie prezentowane są wyniki przewidywań wybranych rezultatów wraz z podsumowaniem skuteczności modeli i ich wzajemnym porównaniem. Zostanie też zwrócona uwaga na możliwość usprawnienia funkcji wyjaśniającej (klasycznych modeli eksplanacyjnych), poprzez wnioski płynące z efektów działania modeli uczenia maszynowego w zakresie wyboru i określenia znaczenia poszczególnych predyktorów oraz doprecyzowania ich wartości. Może to zwiększać efektywność modeli klasycznych i przyczyniać się do poszerzenia teoretycznych podstaw rozumienia zjawisk społecznych, tak w zakresie wyjaśnień globalnych, jak i indywidualnych przypadków, (Biecek 2018). Referat został przygotowany w oparciu o projekt NCBiR Tango ścieżka A (TANGO-IV-A/0048/2019).

Literatura:

Biecek P. (2018). *DALEX: Explainers for Complex Predictive Models in R*. Journal of Machine Learning Research, 19: 1-5.

Breiman, L. (2001). *Random Forests*. Machine Learning, 45: 5–32.. DOI: 10.1023/A:1010933404324

- Cortes, C., Vapnik, V. (1995). *Support-vector networks*. Machine Learning, 20(3): 273–297.
- Domingos, P., Pazzani M. (1997). *On the optimality of the simple Bayesian classifier under zero-one loss*. Machine Learning, 29(2): 103-130.
- Langley P., Iba W., Thompson K. (1992). *An Analysis of Bayesian Classifiers*, Proc. of the 10th National Conference on Artificial Intelligence, Menlo Park, AAAI Press, 1992, pp. 223-228.
- Lewis D. D. (1998). *Naive (Bayes) at forty: The independence assumption in information retrieval*. In *European conference on machine learning*, 4-15. Springer, Berlin, Heidelberg.
- LeDell E., Gill N., Aiello S., Fu A., Candel A., Click C., Kraljevic T., Nykodym T., Aboyou P., Kurka M., Malohlava M. (2023). *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*. R package version 3.40.0.1, <<https://CRAN.R-project.org/package=h2o>>.
- Ho, T.K. (1995). *Random Decision Forest*. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16 August 1995, 278-282.

KAZIMIERZ M. SŁOMCZYŃSKI, INSTYTUT FILOZOFII I SOCJOLOGII PAN

Jakich podatków chcą Polacy? Zmiany w świadomości i predykcje na przyszłość w świetle analizy regresji danych panelowych

Poglądy społeczeństwa na podatki stanowią element legitymizacji polityki państwa względem dochodów ludności i możliwych transferów. Im bardziej postulowane przez ludność wysokości podatków dla różnych grup dochodowych są zgodne z faktycznymi podatkami, tym wyższa legitymizacja systemu i większe poczucie sprawiedliwości. Od 1993 roku respondenci POLPAN pytani są o to, jakie podatki powinni płacić ludzie zarabiający określone kwoty. Przykładowo, w 2018 roku pierwsze pytanie kwestionariuszowe na ten temat brzmiało: Są różne opinie na temat podatków. Ile, Pana (Pani) zdaniem, powinny płacić podatku osoby, które miesięcznie zarabiają brutto 2000 złotych? Kolejne pytania dotyczyły podatków dla miesięcznych zarobków 5000, 12 000 i 60 000 złotych brutto. Respondenci mogli podać postulowaną wysokość podatku w procentach lub w PLN, co w analizach i tak przekładano na procenty.

Odsetek badanych opowiadających się za zerowym podatkiem maleje wraz ze wzrostem dochodów, których podatki mają dotyczyć. W 2018 roku odsetki te wynosiły od 39,5% dla dochodów najniższych do 0,9% dla dochodów najwyższych. Ponieważ od 1993 roku systematycznie rośnie procent osób, które postulują podatki zerowe dla najbiedniejszych, można oczekiwać, iż obecnie (w 2023 roku) procent ten sięga 50,0% osób w wieku 21-65 lat. Polki i Polacy są za umiarkowaną, ale wyraźną progresją podatkową przebiegającą od około 10% do około 30%. Postulowany przez respondentów podatek dla najbiedniejszych (obecnie poniżej 10%) wynika z tego, iż uwzględniono tylko tych, którzy proponują podatek nie-zerowy; wśród nich część respondentów jest za jednolitym podatkiem dla wszystkich (zwykle około 20%). Aż połowa respondentów proponowała, by podatek dla osób zarabiających między 300 a 720 tys. rocznie wynosił co najmniej 25%. Spośród tych respondentów niektórzy proponowali znacznie wyższy odsetek, tak że średni proponowany przez tę grupę podatek to 36,4%. Dla okresu 2008-2018 obserwujemy dużą stabilność proponowanych podatków w szczególności dla środkowych grup zarobkowych. Dla osób zarabiających w granicach 40 000 – 60 000 PLN rocznie średni proponowany podatek wynosi od 11,3% do 13,8%, a dla osób zarabiających w granicach 78 000 – 144 000 PLN rocznie wynosi on około 20%. Są to procenty nieco wyższe niż aktualnie obowiązujące.

W referacie przedstawiono problemy metodologiczne analiz regresji ze zmiennymi opóźnionymi w czasie. Poza tym w analizach panelowych próbowano ustalić wielkość zróżnicowania między-osobniczego w stosunku do zróżnicowania wewnątrz-osobniczego.

Jak zmiana metody głosowania wpływa na zachowania i wyniki wyborcze: Eksperyment naturalny w budżecie obywatelskim m.st. Warszawy

Budżet partycypacyjny (BP) umożliwia zaangażowanie mieszkańców w alokację środków publicznych. Zgłaszają oni swoje projekty wraz z szacowanymi kosztami ich realizacji, a następnie w drodze głosowania wybierają zwycięski zestaw do sfinansowania w ramach dedykowanej kwoty budżetowej. BP jest przykładem grupowego podejmowania decyzji. Sposób oddawania głosów określa, jak możemy wyrażać swoje indywidualne preferencje, a metoda zliczania oddanych głosów – jak przełożą się one na wynik wyborów.

Czy dopóki preferencje mieszkańców pozostają bez zmian, to możemy spodziewać się tego samego wyniku głosowania? Teoria społecznego wyboru oraz badania empiryczne wskazują, że niekoniecznie. Istotne znaczenie odgrywa bowiem dobór samej metody głosowania. Eksperyment naturalny w ramach stołecznego budżetu obywatelskiego stał się impulsem do zbadania wpływu metody głosowania na indywidualne decyzje wyborców oraz ich społeczny rezultat w kontekście BP. Do V edycji obowiązywało w Warszawie głosowanie typu *knapsack*, gdzie można wybrać dowolną liczbę projektów o łącznym koszcie nieprzekraczającym dedykowanej kwoty. Począwszy od VI edycji wprowadzono głosowanie typu *k-approval*, w którym można poprzeć maksymalnie z góry ustaloną liczbę projektów.

Wykorzystanie naturalnego eksperymentu oferuje unikalną okazję ewaluacji realnego zjawiska. Umożliwia dostęp do całego elektoratu w sytuacji dokonywania prawdziwych wyborów na dużą skalę, opartych na rzeczywistych potrzebach i motywacjach głosujących. Jednocześnie brak kontroli nad zmieniającymi się w trakcie eksperymentu uwarunkowaniami (m.in. inne listy projektów, kwoty budżetowe, poziomy głosowań) stanowi szczególne wyzwanie dla analizy porównawczej badanych edycji. W celu sformułowania precyzyjnych wniosków obserwacje empiryczne uzupełniono o wyniki przeprowadzonego eksperymentu online polegającego na głosowaniu w fikcyjnym BP. Połączenie obserwacji pochodzących z rzeczywistych wyborów z wynikami pomocniczego eksperymentu jest nowatorskim podejściem do badania konsekwencji zmiany metody głosowania w BP.

W wystąpieniu omówione zostaną założenia i techniki zastosowane do wnioskowania w oparciu o eksperyment naturalny, projekt dopełniającego eksperymentu online oraz następujące wyniki. Liczba popieraných projektów istotnie różni się dla głosowania *knapsack* w porównaniu do *k-approval*, a wyborcy są bardziej oszczędni w wyborze projektów w ramach tego pierwszego. Skłonność do wspierania tańszych projektów pojawia się tam, gdzie istnieje konieczność uwzględniania ich kosztów. Różne indywidualne strategie głosowania prowadzą do wyboru innych zestawów zwycięskich projektów. W szczególności metoda *knapsack* faworyzuje te tańsze, dlatego więcej projektów może zostać zrealizowanych. Modyfikacja samego sposobu zliczania głosów również wpływa na wyniki wyborów. Zastąpienie powszechnie stosowanego *algorytmu zachłannego* jego wersją uwzględniającą proporcję otrzymanych głosów do kosztów, skutkuje wyborem większej liczby projektów o niższych kosztach. Według opinii ankietowanych głosowanie *knapsack* wymaga większego wysiłku, lecz jednocześnie byłoby chętniej widziane w oficjalnym budżecie. Ponadto, przy wystarczająco długich listach, wyższa pozycja projektu prowadzi do jego większej popularności.

*Projekt badawczy realizowany w ramach stypendium m.st. Warszawy dla doktorantów
CKS/B/V/8/8/U/247/2019