

Paweł Kleka
Instytut Psychologii, UAM

Paradane z kwestionariuszy jako źródło wartościowej informacji o osobach badanych

Wprowadzenie

W dzisiejszych czasach metody kwestionariuszowe, stosowane w badaniach psychologicznych i szerzej społecznych, coraz częściej przybierają formę elektroniczną, a to pociąga za sobą nowe możliwości pomiarowe. Podczas wypełniania testu bądź ankiety komputer bez udziału użytkownika (badacza) i świadomości osoby badanej może zbierać poprzez skrypty CGI, aplety Javy, logi po stronie serwera i użytkownika różne informacje. Te informacje – zwane paradanymi – dotyczyć mogą m.in. mijającego czasu badania. Dane na temat czasu zapisują np. systemy do wywiadów telefonicznych, gdzie pomiar oparty jest na okresach ciszy między pytaniami ankietera a odpowiedziami respondentów. Jednak większa łatwość pozyskania detalicznych danych o przebiegu badania (godzina, czas poświęcony na badanie oraz na poszczególne pytania) nie idzie w parze z popularnością ich analiz. Dane dotyczące czasu odpowiadania mają mocno skośne rozkłady i są często zanieczyszczone przez wiele źródeł potencjalnych wpływów, co zniechęca badaczy do ich analizowania. Celem tej pracy jest pokazanie możliwości analizy paradanych jako źródła wartościowej informacji w badaniach naukowych.

Zachowanie się osób uczestniczących w badaniach internetowych

Badania online jak żadne inne pozwalają na zbieranie różnorodnych danych podczas badania, ale też jednocześnie pozwalają osobom badanym na różnorodne niekontrolowane przez badacza zachowanie. Czynnikiem

zakłócającym mogą być warunki techniczne – różnice w oprogramowaniu (różne przeglądarki, programy filtrujące, anonimizujące) i w sprzęcie (choćby różne monitory). Przy tych samych zadaniach osoby badane mogą nie otrzymywać identycznych bodźców. Różnice między osobami mogą być też spowodowane różnym poziomem opanowania umiejętności komputerowych.

Przeprowadzając badania za pośrednictwem internetu, badacz staje wobec pytania, kim są osoby uczestniczące w badaniu i jak może kontrolować skład próby. Pytania te są o tyle ważne, iż determinują zarówno trafność zewnętrzną (w jakim zakresie wyniki mogą być generalizowane na populację?), jak i trafność wewnętrzną (czy osoby w grupach badawczej i kontrolnej się różnią?) badania. Populacja internautów jest dobrze opisana (Czapiński i Panek, 2015), a stały dostęp do internetu w Polsce ma ponad 75,8% ludności (stan na koniec 2015 wg GUS-u; w niektórych krajach Europy wskaźnik ten zbliża się do 100% (np. w Norwegii, Szwecji, Islandii); dla porównania w Stanach Zjednoczonych, które są pierwszym krajem pod względem bezwzględnej liczby użytkowników internetu, odsetek ten wynosi ponad 77%). Jednakże użytkownicy internetu w porównaniu do całej populacji są młodsi, lepiej wykształceni, pochodzą z większych miast, a ze zmiennych psychologicznych charakteryzują się wyższym poziomem stresu i ekstrawersji (Batorski, 2005), co stawia pod znakiem zapytania możliwości generalizacji wyników badań przeprowadzonych w internecie. Co więcej, chcąc zbudować losową próbę, napotykamy na dodatkowy problem z ustaleniem operatu losowania: tylko około 80% użytkowników internetu ma adres poczty elektronicznej, a 67% osób sprawdza go tylko raz w tygodniu. Dodatkowo wiele osób posiada kilka adresów, co jeszcze bardziej utrudnia przeprowadzenie doboru losowego do prób badawczych. W tej sytuacji badacze polegają często na rekrutacji poprzez ogłoszenia na stronach www, co zwiększa wpływ efektu ochotniczego. Należy dodać, że różne miejsca (fora, serwisy społecznościowe, własne strony internetowe) przeprowadzania rekrutacji ten efekt minimalizują (Batorski i Olcoń-Kubicka, 2006), ale go nie usuwają.

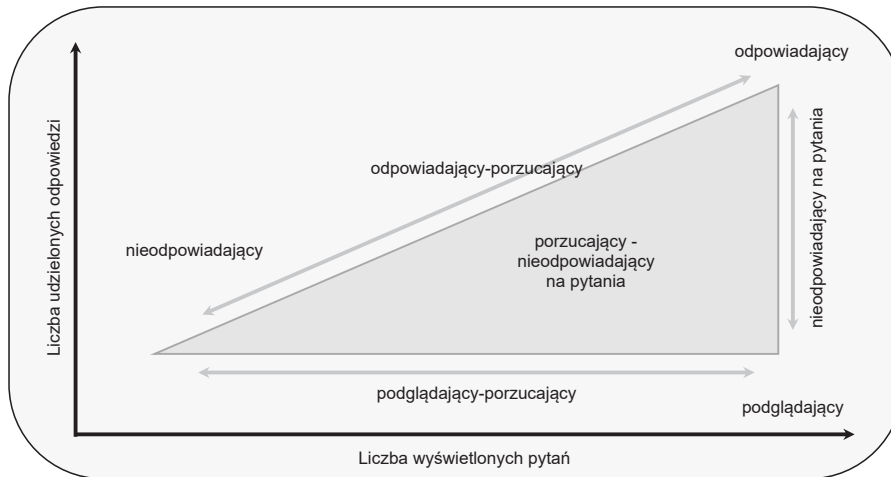
Pewien sposób rozwiązania powyższego problemu stanowi analiza uzyskanych wyników w grupach wydzielonych ze względu na podstawowe zmienne demograficzne: płeć, poziom wykształcenia, wiek i inne. Dzięki dużym próbom takie podejście jest możliwe bez utraty mocy analiz, a uzyskanie zgodnych wyników w poszczególnych warstwach pozwala z większą pewnością i odpowiedzialnością generalizować wyniki na populację (Birnbaum, 2004).

Stopień realizacji próby w badaniach internetowych waha się w okolicach 30% i jest gorszy od badań papierowych (tutaj około 55%) (Nulty, 2008). W większości badań analizuje się dane, które udało się w nich zebrać, a rzadko uwagę poświęca się brakom danych i rezygnacjom z udzielania odpowiedzi. Badacze skupiają się na dwóch aspektach. Po pierwsze, na zwiększeniu odsetka odpowiedzi, ewentualnie skorygowaniu wyników tak, aby uwzględnić osoby, które odmówiły udziału w badaniach (Dillman, Sinclair i Clark, 1993; Bickart i Schmittlein, 1999). Po drugie, na identyfikacji przyczyn rezygnacji z udziału w badaniach (Groves, Cialdini i Couper, 1992; Couper i Rowe, 1996). Pozostaje jednak bez odpowiedzi pytanie, czy czynniki, które spowodowały rezygnację osób zaproszonych do badań, nie mają przypadkiem charakteru stałej cechy wpływającej na wyniki. Innymi słowy, pamiętając o obecności wpływu ochotniczego (szczególnie przy badaniach otwartych), należy także rozpatrzyć powody, dla których osoby badane nie odpowiedziały albo odpowiedziały jedynie częściowo na pytania kwestionariusza.

Pomoc w tym zakresie może stanowić klasyfikacja zachowań osób badanych Michaela Bosnjaka i Tracy Tuten (2001) bazująca na dwóch wymiarach: liczbie zaprezentowanych pytań i liczbie udzielonych odpowiedzi¹. W oparciu o te liczby wyróżniono siedem podgrup osób uczestniczących w badaniach (ryc. 1.). Z punktu widzenia badacza najbardziej pożądana jest grupa osób, które zapoznały się ze wszystkimi pytaniami oraz udzieliły na nie odpowiedzi (*Complete Responders*). Możliwa jest też sytuacja przeciwna, gdy osoby zrezygnują z badania, zanim się ono rozpocznie (*Nonresponders*). Powodem mogą być przeszkody techniczne bądź decyzja o rezygnacji, np. po zapoznaniu się z planszą tytułową, co odbywa się bez kontaktu z pozycjami narzędzia. Trzecią grupę stanowią obserwatorzy. To osoby, które chcą zapoznać się z metodą, ale nie biorą udziału w badaniu (*Lurkers*) – przeglądają one wszystkie pozycje, ale nie udzielają żadnych odpowiedzi. Osoby, które zapoznając się z kolejnymi pozycjami, zrezygnowały z udziału w badaniu, w zależności od tego, czy udzielały odpowiedzi, klasyfikowane są albo jako odpowiadający-porzucający (*Answering Drop-Outs*), albo jako obserwatorzy-porzucający (*Lurking Drop-Outs*). Ostatnie dwie grupy stanowią osoby, które zapoznały się z częścią bądź całością

¹ Autorzy klasyfikacji przyjęli trzy założenia: osoba uczestnicząca w badaniu za pośrednictwem internetu może zobaczyć wszystkie pozycje, które pojedynczo wyświetlane są na monitorze; może przejść do dalszych pozycji, nie odpowiadając na poprzednie; jej kolejne odpowiedzi są zapisywane.

pozycji w badaniu i udzieliły odpowiedzi na niektóre z nich. Są to osoby, które zrezygnowały przed końcem badania, określane mianem porzucających – nieodpowiadających na pytania (*Item Nonresponding Drop-Outs*) lub rezygnujące po zapoznaniu się z całą metodą – nieodpowiadające na pytania (*Item Nonresponders*).



Ryc. 1. Klasyfikacja typów zachowań osób uczestniczących w badaniu ze względu na relację między liczbą pytań, z którymi osoba się zapoznana, a liczbą udzielonych odpowiedzi

Źródło: Bosniak i Tuten (2001).

Oczywiście najbardziej pożądanym jest udział osób odpowiadających na wszystkie pozycje, z minimalnym udziałem grup osób, które porzucają badanie. Według wspomnianych autorów umożliwiają to dwa kierunki działań: 1) dostosowywanie formy badania do oczekiwań osób uczestniczących, minimalizując ewentualną frustrację, która sprzyja porzuceniu badania oraz 2) motywowanie do pozostawania w badaniu.

Badanie internetowe zapewnia respondentom wysokie poczucie własnej anonimowości, mimo możliwej identyfikacji komputera (jest to informacja, która powinna znaleźć się w prośbie o zgodę na badanie). Ponieważ osoby uczestniczące w badaniu czują się anonimowo i nie ponoszą wysokich kosztów psychicznych w przypadku rezygnacji, to oddziaływanie na ich motywację do pozostawania w badaniu zmieniające bilans zysków i strat przynosi dobre efekty. Ujęcie obejmujące najważniejsze czynniki motywacyjne zaproponował Ulf-Detrich Reips w artykule *Standards for*

Internet-based experimenting (2001). Jego metoda tzw. wysokiego progu (*high-hurdle*) zakłada, że w planowaniu badania internetowego należy zadbac o następujące czynniki wpływające na motywację do jego ukończenia:

- Powaga – poprzez poinformowanie uczestników badania o ważności badań i potrzebie uzyskania rzetelnych wyników.
- Personalizacja – poprzez zapytanie o adres mailowy lub numer telefonu.
- Przeświadczenie o kontroli – poprzez poinformowanie o identyfikowaniu komputera, z którego ankieta jest wypełniana.
- Cierpliwość – najwięcej treści należy zamieścić na początku narzędzia, sukcesywnie zmniejszając ich ilość ze strony na stronę.
- Czas trwania – trzeba podać orientacyjny czas trwania badania.
- Prywatność – na początku badania poinformować osoby, które zechcą uczestniczyć w badaniach, że zostaną zadane pytania, np. o poziom dochodów.
- Warunki – należy określić wymagane elementy oprogramowania i podać odnośniki do ich pobrania i zainstalowania.
- Pretest techniczny – przeprowadzenie testu sprawdzającego, czy komputer osoby zaproszonej do wzięcia udziału w badaniach spełnia wymagania techniczne.
- Nagroda – informacja o nagrodzie przyznanej / wylosowanej za pełny udział w badaniu.

Dodatkowo, według Reipsa, poprzedzając właściwą fazę badania kilkoma „pytaniami na rozgrzewkę” (*warm-up phase*), zwiększa się prawdopodobieństwo zebrania rzetelnych danych, dzięki zaznajomieniu osób badanych ze sposobem odpowiadania w danym kwestionariuszu.

Proces odpowiadania na pytania w badaniach kwestionariuszowych

Osoba uczestnicząca w badaniu nie jest w homeostazie przez cały czas jego trwania. Im dłuższy czas badania, tym większe prawdopodobieństwo spadku motywacji i wzrostu zmęczenia przy udzielaniu odpowiedzi (Bieček, 2016). Podczas udzielania odpowiedzi na kolejne pytania zachodzą w osobie badanej też różnorodne procesy poznawcze (fazy przetwarzania – opisane w dalszej części), które mają wpływ na odpowiedzi (Jabine et al., 1984; Loftus, Fienberg i Tanur, 1985; Tourangeau, Rips i Rasinski, 2000; Schwarz, 2007).

Przyjmując perspektywę poznawczą, proces udzielania odpowiedzi można podzielić na następujące fazy: przetwarzanie informacji zawartych w pytaniu prowadzące do zrozumienia, przywoływanie informacji, faktów i zdarzeń z pamięci, konstruowanie odpowiedzi na podstawie posiadanych informacji i wreszcie kodowanie ich tak, aby pasowały do wymaganego formatu odpowiedzi. Każda z tych faz może być źródłem specyficznych błędów (Tourangeau et al., 2000, s. 7). Skuteczne przejście przez te procesy (uwarunkowane zarówno możliwościami poznawczymi, jak i stanami emocjonalnymi lub tendencjami osobowościowymi) sprzyja udzieleniu kompletu odpowiedzi. Zakłócenia natomiast są odpowiedzialne za pozostałe formy odpowiadania.

Faza pierwsza, przetwarzanie informacji zawartych w pytaniu, prowadzi do zrozumienia pytania – zainteresowanie się pozycją kwestionariuszową, prześledzenie instrukcji, rozpoznanie sensu polecenia. Zauważa się, że inaczej przetwarzane są pytania dotyczące postaw (badany może odwołać się do wcześniejszych ustaleń lub sformułować osąd od zera), a inaczej pytania o zachowanie (badany musi zidentyfikować potrzebne zachowanie i przywołać wydarzenia z pamięci autobiograficznej). W tej fazie mogą mieć źródło problemy skutkujące błędnymi odpowiedziami, jeśli na przykład osoby uczestniczące w badaniu nie rozumieją pytania lub rozumieją je niezgodnie z intencją badacza. Problemy te mogą dotyczyć uwagi i skupienia się osoby uczestniczącej w badaniu lub być wywołane przez nieprzyjazne sformułowanie pozycji (zdania złożone, niedopasowane językowo, wieloznaczne) przez twórcę narzędzia. Druga faza polega na przywoływaniu informacji z pamięci długotrwałej według wybranej strategii i/lub wypełnianiu luk w pamięci. Jeśli osoba badana nie może przywołać wystarczającej lub niezbędnej do udzielenia odpowiedzi informacji, może dokonać interpretacji pozycji kwestionariuszowej według własnego uznania, co nie zawsze musi się pokrywać z intencją badania. W trzeciej fazie, polegającej na sformułowaniu własnej opinii, zdarza się, że przywołane z pamięci informacje są niewystarczające i wymagają dodatkowego przetwarzania (np. pytanie o liczbę wizyt w zadanym okresie wymaga oprócz przypomnienia wizyt zsumowania ich). Faza ta obejmuje wszystkie procesy potrzebne do połączenia lub uzupełnienia informacji – także te, gdy wskutek braków przywoływane są opinie obiegowe lub formowane ad hoc na bazie różnych heurystyk. Ostatnia faza obejmuje procesy poznawcze dotyczące dopasowania wygenerowanej informacji do kafeterii odpowiedzi przygotowanych przez badacza. Osoba uczestnicząca w badaniu dokonuje wyboru odpowiedzi, biorąc pod uwagę i jej treść, i aspekty subiektywne (np. dokonu-

je rozróżnienia pomiędzy odpowiedziami „zdecydowanie tak“ oraz „tak“). Obserwowalne wydłużenie czasu reakcji na pozycję kwestionariuszową (efekt odpowiedzi, *response effect*) może dotyczyć dowolnej z tych faz – być spowodowane niezrozumieniem pytania, problemami z przypomnieniem informacji lub trudnościami ze znajdowaniem właściwej opcji odpowiedzi. Na podstawie czasu trwania efektu odpowiedzi i w jego relacji do rodzaju i treści pytania można pokusić się o stawianie hipotez na temat przebiegu procesów wewnętrznych u danej osoby analizując jej czasy odpowiadania, a także na temat interakcji między pozycjami testowymi a osobą uczestniczącą w badaniu, przez analizę czasów odpowiadania, jakie charakteryzowały poszczególne pytania. Obserwując ponadprzeciętny czas odpowiedzi, można zastanowić się, co w pytaniu sprawiło trudność. Identyfikując zwiększoną liczbę zmian odpowiedzi – czy pozycja była niejasna, a przy zmianach kolejności odpowiadania („przeskakiwaniu” między pozycjami) – czy była drażliwa. Z kolei pytania dotyczące wpływu pozycji mogą być następujące: Czy wszystkie osoby reagowały na nie w podobny sposób? Ile osób odpowiedziało po wymuszeniu odpowiedzi przez kwestionariusz (tzn. w sytuacji, gdy brak odpowiedzi uniemożliwiał przejście do kolejnego kroku), a ile osób zrezygnowało z dalszego udziału w badaniu? Na te i podobne pytania można próbować odnaleźć odpowiedź w paradanych, dlatego wzrasta wśród badaczy przekonanie o zaletach analizy czasu odpowiadania i jego charakterystyki jako predyktora zachowań podczas badania (Schwarz, 2007; Yan i Tourangeau, 2008). W skrajnych przypadkach niektórzy badacze doszli nawet do wniosku, że czas odpowiedzi jest lepszym predyktorem rzeczywistych zachowań niż sama odpowiedź (Bassili i Scott, 1996).

Analiza czasu odpowiadania

Potencjał analiz czasu odpowiadania rozwija się wraz z możliwościami technologii pomiarowych od: 1) czasu reakcji na zaproszenie do badań; przez 2) czas odpowiedzi na poszczególne pytania jako miary jakości odpowiedzi; 3) czasu odpowiedzi jako funkcji trudności pytań; aż do 4) klasyfikacji zachowań osób badanych.

Zainteresowanie czasem wykonywania zadań towarzyszyło psychologom od początku kwestionariuszowego badania postaw, przekonań i zachowań człowieka, ale ze względu na ograniczenia techniczne pierwsze analizy dotyczyły jedynie czasu reakcji na zaproszenie do badania (Rao i Pennington, 2013). W tym obszarze autorzy wykazali, że od osób, które

odpowiedziały na zaproszenie, stosunkowo szybko można spodziewać się większej motywacji i zaangażowania, mniej też u tych osób potem braków danych i błędnych odpowiedzi. Stwierdzono też istotne różnice demograficzne między tą grupą (mężczyźni, w wieku 55+, właściciele mieszkania/domu, często używający internetu) a osobami, które przystępują do badań pod koniec terminu lub pod wpływem kolejnego przypomnienia (Rao i Pennington, 2013, s. 669).

Kolejny sposób uwzględniania czasu w badaniach polegał na ocenie jakości danych w zależności od czasu odpowiedzi (Greszki, Meyer i Schoen, 2015). Dzięki wyznaczeniu rozkładów czasów odpowiedzi na poszczególne pytania stało się możliwe zdefiniowanie tych obserwacji, które ze względu na zbyt krótki czas odpowiedzi rodziły podejrzenie o nierzetelnym wypełnianiu przez osoby badane. Warto wspomnieć, że w przytoczonych badaniach usuwanie takich nierzetelnych danych lub całych wektorów odpowiedzi (obserwacji) nie wykazało żadnej znaczącej różnicy w wynikach, a jedynie zmniejszało poziom błędu w danych. Mimo że dane nierzetelne nie różniły się od pozostałych, to zabieg tego rodzaju pozwolił zredukować ilość szumu w zebranych danych (Rao i Pennington, 2013, s. 498).

Rola czasu w ocenie jakości odpowiedzi została inaczej potraktowana przez Ting Yana i Rogera Tourangeau (2008). Przenieśli oni akcent z traktowania czasu jako wskaźnika wewnętrznej dostępności pojęć u osoby badanej na analizę czasu jako składowej charakterystyki pozycji testowej i stylu odpowiadania respondentów. Pozwoliło to poddać analizie sposób odpowiadania, gdzie większość wariacji czasu odpowiadania tłumaczona jest przez proces zaznajamiania się (tu: czytania) z pozycją testową (istotne są cechy takie jak długość oraz liczba słów), ale istotne części wariacji czasu odpowiadania tłumaczone są także przez rodzaj i liczbę kategorii odpowiedzi, umiejscowienie pytania w kwestionariuszu oraz właściwości osoby badanej (tu: wiek, wykształcenie, doświadczenie w korzystaniu z internetu). Podobne wyniki uzyskali Mike Couper i Franke Kreuter (2013), wykazując, że różnice w czasie rozwiązywania zadań zależą od długości i rodzaju zadań testowych oraz od ilości kategorii odpowiedzi.

Chun Wang i Gongjun Xu poszli o krok dalej i dzięki zastosowaniu zaawansowanych metod statystycznych zaproponowali model, który pozwala na rozróżnienie różnych rodzajów zachowań osób badanych (Wang i Xu, 2015; Wang, Xu i Shang, 2016). Przywołani badacze wykazali, że w oparciu o mieszane modele hierarchiczne oraz indeks resztowy w paradygmacie Bayesowskim (*Bayesian residual index*) możliwe jest na podstawie czasu odpowiedzi zidentyfikowanie różnych zachowań. Udało im się rozróżnić

rozwiązywanie zadań (*solution*), oszukiwanie (*cheating*) oraz zgadywanie (*rapid guessing*) (Wang et al., 2016, s. 21). Przeprowadzenie takiej klasyfikacji jest łatwiejsze w testach z prawidłową odpowiedzią, gdzie progiem detekcji jest zbyt krótki czas odpowiadania przy poprawnych odpowiedziach. Przy czym im trudniejsze jest pytanie, tym łatwiejsze jest wykazanie zgadywania (lub oszukiwania), ponieważ różnica czasu między rozwiązywaniem a „złym“ zachowaniem jest większa (Marianti et al., 2014).

Zgadywaniem odpowiedzi przez osoby badane zajmują się również modele IRT, które zakładają, że prawdopodobieństwo poprawnej (zgodnej z kluczem) odpowiedzi rośnie wg funkcji logitowej wraz ze wzrostem poziomu badanej cechy (Θ), zaś poziom zgadywania jest stały dla danej pozycji testowej. Analiza czasów odpowiedzi pozwala zaproponować alternatywne wyjaśnienie – zgadywanie jest zależne od kontekstu i pojawia się wtedy, gdy osoba badana jest zmotywowana do udzielenia odpowiedzi bez przetwarzania pytania. Dzieje się tak przeważnie w sytuacji ograniczonego czasu lub niskiego kosztu udzielenia błędnej odpowiedzi. Przy takim wyjaśnieniu zjawisko zgadywania jest mniej, jeśli w ogóle, powiązane z budową pytania (Wise i Kong, 2005).

Modelowanie statystyczne czasu odpowiadania

Analiza czasów odpowiadania pozwala na wgląd w przebieg procesu odpowiadania na kwestionariusz – np. możliwe staje się zidentyfikowanie sytuacji, gdy osoby badane zgadują odpowiedzi na końcowe pytania kwestionariusza ze względu na kończący się czas w tzw. testach szybkości (*speed tests*) (Bolt, Cohen i Wollack, 2002). W analizach czasu odpowiedzi jako zmiennej zależnej możemy wyróżnić dwie grupy modeli:

1. Bazujące na rozkładzie czasu niezależnie od charakterystyki pozycji i osób badanych. Innymi słowy czasy odpowiadania są modelowane niezależnie od poprawności odpowiedzi (Maris, 1993; Schnipke i Scrams, 1997).
2. Opisujące rozkład czasów odpowiadania w kontekście poprawności odpowiedzi – czas odpowiadania i mierzone cechy są ze sobą w związku (Marianti et al., 2014).

Te pierwsze modele częstsze są w badaniach psychologicznych (w większości przypadków pozycje testowe posiadają tylko klucz odpowiedzi zgodnych z badaną cechą), te drugie w edukacyjnych (klucz odpowiedzi zawiera informacje o odpowiedziach prawidłowych).

Analiza czasu odpowiedzi nastęrcza problemy statystyczne – odpowiedzi na poszczególne pytania są zagnieżdżone w odpowiedziach danej osoby na wszystkie pytania kwestionariusza. Rozwiązaniem jest zaproponowane przez Wima van der Lindena (2007) modelowanie hierarchiczne wprowadzające dwie zmienne latentne: 1) osoby badane oraz 2) pozycje testowe, które wyjaśniają oba procesy (czas odpowiadania i poprawność). Proces udzielania odpowiedzi charakteryzowany w hierarchicznym modelu mieszanym pozwala na identyfikacje zależności wewnątrzobiektywnych, międzyobiektywnych oraz krzyżowych (interakcji między osobami a pozycjami testowymi).

Bez względu na przyjęte założenia przy modelowaniu czasów odpowiedzi pojawia się problem dużej prawoskośności danych. Badacze proponują remedium w postaci transformacji Box-Coxa w miejsce słabszej transformacji logarytmicznej (Klein-Entink, van der Linden i Fox, 2009). Dzięki temu analiza zachowań może być przeprowadzona za pomocą metod parametrycznych i opierać się na znanej analizie reszt i wartości odstających.

Przykład analizy jakościowej

Aby skorzystać z informacji „ukrytej“ w czasach odpowiedzi, można też posłużyć się metodami jakościowymi. Prezentując dane zebrane przez systemy komputerowe (milisekundy) w postaci czytelnej dla człowieka (godziny, minuty, sekundy), można wnioskować o rodzaju zachowania osób uczestniczących w badaniach. W celu ilustracji takiej analizy przygotowano dane z badania postaw studentów psychologii za pośrednictwem strony internetowej (badanet.amu.edu.pl), przeprowadzonego w 2013 r. (tab. 1.). Wyniki badań nie były analizowane dotąd pod względem RT. Na potrzeby niniejszej pracy wykorzystano tylko fragment wyników.

Analiza godziny udzielenia odpowiedzi oraz odstępu między kolejnymi odpowiedziami pozwala stawiać różnorakie hipotezy na temat przebiegu badania. Jeśli zaobserwowano długi czas między wejściem na stronę a udzieleniem pierwszej odpowiedzi, można sądzić, że osoby uczestniczące w analizowanym badaniu skłonne były odłożyć udzielanie odpowiedzi np. do następnego dnia. Można to tłumaczyć tym, że zaproszenie do badań rozsyłane było za pośrednictwem poczty elektronicznej i osoby po przeczytaniu maila i kliknięciu w link prowadzący do strony z pytaniami zostawiały sobie to zadanie na bardziej sprzyjającą porę. Dla niektórych osób był to następny ranek, ale były też osoby, które wolały dokończyć odpowiadanie w późnych godzinach nocnych.

Tabela 1. Godzina i czas udzielenia odpowiedzi na poszczególne pytania w kwestionariuszu. Fragment danych

Start	Pytanie 1	Pytanie 2	Pytanie 3	Pytanie 4	Pytanie 5	Zachowanie Osoby badanej
08:17:47	08:21:14	09:34:52	10:59:14	15:21:13	15:21:16	na raty
08:18:09	13:37:43	21:49:32	21:52:34	21:53:48	21:58:13	na raty
08:18:33	17:35:40	17:10:36	17:11:39	17:12:01	17:14:39	poprawione pytanie nr 2
08:18:40	21:26:42	21:43:36	21:44:16	21:44:55	21:38:31	w nocy
09:04:31	22:06:53	22:15:00	22:16:40	22:17:32	22:20:40	w nocy
09:08:14	22:39:33	22:56:04	22:59:05	22:59:58	23:01:37	w nocy
09:47:22	23:21:39	23:37:51	23:38:48	23:39:13	23:40:37	w nocy
09:49:10	00:22:27	03:33:17	03:34:45	03:35:28	03:36:48	na raty, nad ranem
09:57:56	06:38:17	07:13:04	07:15:16	07:16:11	07:19:55	następnego dnia
10:00:04	07:22:14	07:32:21	10:23:34	10:37:42	10:38:37	na raty, następnego dnia
10:03:35	07:46:13	07:57:50	08:09:54	08:10:55	08:14:54	następnego dnia
10:03:44						odmowa
10:41:28	08:24:16	10:54:43	10:54:59	10:55:01		porzucenie (lurker)
08:23:48						odmowa
10:42:18	08:37:28	08:49:33	08:50:58	08:52:02	08:54:21	następnego dnia
10:42:46	08:52:55	09:06:07	09:11:39	09:12:22	09:14:11	następnego dnia
10:43:25	09:12:40					porzucenie (drop-out)
11:43:05	11:43:41	11:43:43	11:38:42	11:38:46	11:38:54	jednokrotnie
10:45:16	09:49:16	10:02:53	10:09:15	10:09:36	10:10:25	następnego dnia
10:15:38	10:16:13	10:28:26	10:42:41	10:43:38	10:47:03	jednokrotnie
10:54:00	10:51:02	11:44:59				porzucenie (drop-out)
10:57:07	11:31:50	11:53:13	12:10:04	12:10:36	12:12:17	na raty
Przeciętny czas odpowiadania (<i>response time</i>) na poszczególne pytania						
Q25	8 sek.	1,88 min	55 sek.	27 sek.	53 sek.	
Mediana	11 sek.	3,15 min	1,22 min	34 sek.	1,77 min	
Q75	19 sek.	5,23 min	1,73 min	34 sek.	1,47 min	

Źródło: badanie własne, N = 49, kwestionariusz zawierał zróżnicowane pytania (1, 4 – jednokrotnego wyboru, 2 – otwarte, 3, 5 – wielokrotnego wyboru). Surowe dane zebrane w milisekundach przeliczono na prostszą postać w celu ułatwienia prezentacji.

Oprócz pory dnia, w której osoby udzieliły odpowiedzi, na podstawie długości odstępów między poszczególnymi czasami, można wnioskować, czy wypełnianie kwestionariusza przebiegało w sposób płynny, czy też było rozłożone na raty. Jeśli odstęp między poszczególnymi odpowiedziami przekraczał przeciętną długość odpowiedzi na dane pytanie, to osoba badana przerwała badanie i wróciła doń po dłuższej (mierzalnej) przerwie.

Rejestracja czasu pozwala także zidentyfikować te przypadki, gdzie osoba cofnęła się w kwestionariuszu, aby zmienić swoją odpowiedź (godziny odpowiedzi na kolejne pytania nie następują po sobie).

Za pomocą analizy czasu możliwe też staje się rozróżnienie osób porzucających badanie na takie, które odpowiadały, lecz coś przeszkodziło im w ukończeniu badania (*drop-out*), od takich, które udzieliły odpowiedzi bez znaczenia i chciały tylko zapoznać się z treścią pytań (*lurker*). Tutaj pomocnym punktem odniesienia może być mediana czasów odpowiedzi wraz z pierwszym i trzecim kwartyłem, pokazujące typowość czasów odpowiadania dla poszczególnych pytań.

Po dokonaniu jakościowej klasyfikacji sposobów zachowania badacz ma możliwość zweryfikowania ich wpływu na wyniki przez zbadanie różnic między grupowych. Jeśli treść kwestionariusza jest wrażliwa na styl odpowiadania, może to wpłynąć na decyzję o innym sposobie zbierania danych lub zmianie samego narzędzia badawczego (np. przez ograniczenie pory dostępu, wymuszenie jednokrotnego rozwiązywania, rozszerzenie instrukcji o standaryzację warunków badanie).

Podsumowanie

Czas udzielania odpowiedzi przez respondentów był od dawna przedmiotem badań w psychologii, jednakże systematyczne uwzględnienie go w kwestionariuszowych badaniach online jest jeszcze rzadkością. Dane o czasie odpowiadania okazują się użyteczne na kilka sposobów.

Pierwszym byłyby ocena jakości zebranych odpowiedzi. Przyjmując zaprezentowany w tej pracy poznawczy model udzielania odpowiedzi, należy podkreślić, że nawet w taekwiwalentnych kwestionariuszach (składających się z homogenicznych pozycji) pozycje testowe różnią się od siebie ze względu na to, jak są odczytywane i rozumiane. Wydaje się, że rozszerzenie modeli powiązań między zmiennymi o czynnik czasu i uwzględnienie procesów przetwarzania podczas odpowiadania pozwoli lepiej wytłumaczyć zmienność intraindywidualną osób badanych. Dodatkowo rozpoznanie

sposobów odpowiadania na pozycje kwestionariusza w oparciu o analizę czasu (nawet tak proste, jak zaprezentowane w tej pracy) zmniejsza wadę badań online związaną z brakiem kontroli nad przebiegiem badania.

Drugim aspektem byłaby ocena trafności narzędzi badawczych. Potraktowanie czasu jako nowej zależnej zmiennej, w miejsce przyjmowania go tylko za kowariant lub wskaźnik badanych cech, dostarcza danych do oceny trafności teoretycznej kwestionariusza. Szczegółowa analiza czasów pozwala badaczom uzyskać informacje na temat samego narzędzia. Wykrycie nieregularności w czasach odpowiadania osób badanych pozwala wykryć błędy w jego konstrukcji, natomiast wiedza, czy poprawne odpowiedzi pochodzą ze zgadywania, czy oszukiwania pozwala dokonać niezbędnych korekt w budowie pozycji kwestionariuszowych lub procedurze przeprowadzania badania.

W badaniach tradycyjnych (papier i ołówek), jak i w badaniach nowoczesnych (internetowych lub szerzej elektronicznych) problemem pozostaje detekcja osób udzielających nierzetelnych odpowiedzi. Istniejące metody weryfikacji (porównaj np. skale kontrolne w MMPI-2 (Brzezińska i Koć-Januchta, 2012) obciążają dodatkowymi pytaniami osoby uczestniczące w badaniu i mimo starań badaczy interferują z treścią pytań / zadań właściwych. W tym kontekście analiza czasu, jaki jest poświęcany na poszczególne elementy kwestionariusza przez osobę badaną, może dostarczyć podobnych informacji bez wprowadzania dodatkowych źródeł wariancji.

Literatura

- Bassili, J., Scott, S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, 60, 390–399.
- Batorski, D. (2005). Społeczne aspekty korzystania z nowych technologii. W: J. Czapiński, T. Panek (red.), *Diagnoza społeczna 2005: Warunki i jakość życia Polaków* (s. 27–28, 214–231). Warszawa: Wyższa Szkoła Finansów i Zarządzania.
- Batorski, D., Olcoń-Kubicka, M. (2006). Prowadzenie badań przez internet – podstawowe zagadnienia metodologiczne. *Studia Socjologiczne*, 182, 99–132.
- Bickart, B., Schmittlein, D. (1999). The distribution of survey contact and participation in the United States: Constructing a survey-based estimate. *Journal of Marketing Research*, 36, 286–294.
- Biecek, P. (12.10.2016–14.10.2016). How to use R to hack the publicly available data about skills of 2M+ worldwide students? Paper presented at the eRum!, Poznań.
- Birnbaum, M.H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, 55, 803–832. doi: 10.1146/annurev.psych.55.090902.141601

- Bolt, D.M., Cohen, A.S., Wollack, J.A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement*, 39, 331–348.
- Bosnjak, M., Tuten, T.L. (2001). Classifying response behaviors in web-based surveys. *Journal of Computer-Mediated Communication*, 6.
- Brzezińska, U., Koć-Januchta, M. (2012). Normalizacja. W: J. Stańczak (red.), *Minnesocki Wielowymiarowy Inwentarz Osobowości-2 - MMPI-2*. Warszawa: Pracownia Testów Psychologicznych.
- Couper, M., Kreuter, F. (2013). Using paradata to explore item level response times in surveys. *Journal of the Royal Statistical Society*, 176, 271–286.
- Couper, M., Rowe, B. (1996). Evaluation of a computer-assisted self-interview component in a computer-assisted personal interview survey. *Public Opinion Quarterly*, 60, 89–105.
- Czapiński, J., Panek, T. (2015). *Diagnoza społeczna 2015: Warunki i jakość życia Polaków*. Warszawa: Wyższa Szkoła Finansów i Zarządzania.
- Dillman, D.A., Sinclair, M.D., Clark, J.R. (1993). Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys. *Public Opinion Quarterly*, 57, 289–304.
- Greszki, R., Meyer, M., Schoen, H. (2015). Exploring The Effects of Removing “too fast” Responses and respondent from web surveys. *Public Opinion Quarterly*, 79 (2). doi: 10.1093/poq/nfu058.
- Groves, R.M., Cialdini, R.B., Couper, M. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475–495.
- Jabine, T.B., Straf, M. L., Tanur, J.M., Tourangeau, R. (1984). Cognitive aspects of survey methodology: Building a bridge between disciplines. Washington: National Academy Press.
- Klein Entink, R.H., van der Linden, W.J., Fox, J.-P. (2009). A Box-Cox normal model for response times. *British Journal of Mathematical and Statistical Psychology*, 62, 621–640. doi: 10.1348/000711008X374126.
- Loftus, E.F., Fienberg, S. E., Tanur, J. M. (1985). Cognitive psychology meets the national survey. *American psychologist*, 40, 175–180.
- Marianti, S., Fox, J.-P., Avetisyan, M., Veldkamp, B.P., Tijmstra, J. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics*, 39 (6), 426–451. doi: 10.3102/1076998614559412.
- Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and application as psychometric models for response times. *Psychometrika*, 58 (3), 445–469.
- Nulty, D.D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, 33 (3), 301–314. doi: 10.1080/02602930701293231.
- Rao, K., Pennington, J. (2013). Should the third reminder be sent? The role of survey response timing on web survey results. *International Journal of Market Research Society*, 55(3), 651–674. doi: 10.2501/IJMR-2013-056.

- Reips, U.-D. (2001). Standards for Internet-based experimenting. *Experimental Psychology*, 49, 243–256.
- Schnipke, D.L., Scrams, D.J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, 34, 213–232.
- Schwarz, N. (2007). Cognitive Aspects of Survey Methodology. *Applied Cognitive Psychology*, 21, 277–287. doi: 10.1002/acp.1340.
- Tourangeau, R., Rips, L.J., Rasinski, K. (2000). *The Psychology of Survey Response*. London: Cambridge University Press.
- van der Linden, W.J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72, 287–308.
- Wang, C., Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 68, 456–477. doi: 10.1111/bmsp.12054.
- Wang, C., Xu, G., Shang, Z. (2016). A two-stage approach to differentiating normal and aberrant behavior in computer based testing. *Psychometrika*, 81 (327), 1–32. doi: 10.1007/s11336-016-9525-x.
- Wise, S.L., Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18 (163–183). doi: 10.1207/ s15324818ame1802_2.
- Yan, T., Tourangeau, R. (2008). Fast Times and Easy Questions. The Effects of Age, Experience and Question Complexity on Web Survey Response Times. *Applied Cognitive Psychology*, 22, 51–68. doi: 10.1002/acp.1331.