

Konrad Gałuszko

ORCID: 0000-0003-1298-7765

Joanna Lewczuk

ORCID: 0000-0003-4812-6537

Konrad Krystian Kuźma

ORCID: 0000-0001-7159-6903

Walidować? Weryfikować? Nie ruszać? O niestatystycznych, statystycznych i stochastycznych metodach oceny jakości danych ilościowych opowieść

SŁOWA KLUCZOWE:

*ocena jakości danych, metody statystyczne, kontrola logiczna,
testy statystyczne, rozkład normalny*

STUDIA I ANALIZY

Wstęp

Celem artykułu jest zaprezentowanie możliwych metod weryfikacji danych ilościowych z wykorzystaniem narzędzi niestatystycznych, statystycznych oraz stochastycznych. I choć najprostszymi z nich są metody zdroworozsądkowe, oparte na pewnych założeniach logicznych, to jednak nie mają one umocowania w matematyce, a co za tym idzie – ich wyniki mogą być podważane przez naszych adwersarzy¹.

Nieco inaczej jest w przypadku metod statystycznych, które swoimi korzeniami sięgają starożytnego Egiptu. Już wtedy przeprowadzano pierwsze spisy powszechne², a ich wyniki prezentowano w wygodnym

¹ Nie oznacza to jednak, że są to metody nieużyteczne, każdą metodę bowiem można spróbować podważyć. Choć jest to trudniejsze w przypadku metod statystycznych i stochastycznych, nadal pozostaje możliwe.

² Zob. D. Valbelle, *Les recensements dans l’Égypte pharaonique des troisième et deuxième millénaires*, „CRIPEL” 1987, nr 9, s. 37–49; P. Cartledge, P. Garnsey, E.S. Gruen (red.), *Hel-*

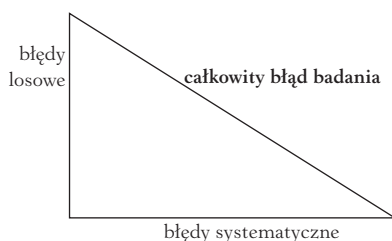
układzie tabelarycznym zawierającym podstawowe informacje o stanie państwa. Do czasów Pierre'a de Fermata oraz Blaise'a Pascala minęło wiele lat. W XVII wieku obaj badacze rozpoczęli analizę gier losowych, co pozwoliło na znaczne rozszerzenie dostępnego katalogu metod statystycznych. Jednak współczesną statystykę, w jej obecnym kształcie, zawdzięczamy w dużej mierze Andriejowi Kołmogorowowi, który przed II wojną światową dokonał aksjomatyzacji teorii prawdopodobieństwa.

Wraz z rozwojem metod statystycznych badacze zaczęli zastanawiać się, w jaki sposób można je ulepszyć. Wynikiem ich działań są metody stochastyczne, które oprócz czynników głównych dają również możliwość – w większym stopniu niż wspomniane wcześniej metody statystyczne – uwzględnienia czynników losowych w celu lepszego poznania badanych zależności³.

Przedstawione w ramach artykułu egzemplifikacje są najbliższe sercu autorów, ale prezentowane metody są pewną prawdą uniwersalną i – jako takie – mogą być zastosowane do większości już istniejących (i nowo tworzonych) zbiorów danych ilościowych. Przedstawiane przez autorów metody są bardzo dobrze rozpoznane, przetestowane i mają bogatą literaturę opisującą ich działania. Jako takie – dysponują również licznymi narzędziami analitycznymi, a także są zaimplementowane w wielu dostępnych programach umożliwiających analizę statystyczną.

Najczęściej ocena jakości danych ilościowych jest częścią problemu związanego z błędami popełnianymi w trakcie realizacji badania. Na błąd całkowity składają się błędy losowe i błędy systematyczne. Relacje między tymi błędami przedstawia rysunek 1.

Rysunek 1. Zależności między elementami składowymi całkowitego błędu badania



Źródło: opracowanie własne na podstawie: G. Lissowski, *Z zagadnień doboru próby*, [w:] K. Szaniawski (red.), *Metody statystyczne w socjologii: wybrane zagadnienia*, Warszawa 1968, s. 68.

lenistic Constructs: Essays in Culture, History, and Historiography, Berkeley – Los Angeles – Londyn 1997, s. 242.

³ F. Bławat, *Podstawy analizy ekonomicznej. Teorie, przykłady, zadania*, Warszawa 2011, s. 35.

Ujęcie tego problemu zaprezentowane na rysunku 1 po raz pierwszy przedstawiono na gruncie polskiej socjologii w dziele, na podstawie którego sporządzono rysunek⁴. Zgodnie z tym ujęciem błędy losowe dotyczą doboru próby i są całkowicie możliwe do obliczenia (skwantyfikowania). Błędy systematyczne mają z kolei dwoistą naturę: mogą wynikać zarówno z procedury wybierania próby (na przykład wadliwego operatu losowania), jak i elementów badania niezwiązanych z samą próbą. Klemens Szaniawski wymienia tu trzy źródła tego podtypu błędu systematycznego:

- występowanie jednostek niedostępnych⁵;
- błąd mierzenia – chodzi tu głównie o błędy odpowiedzi (*errors in response*), których źródłem może być respondent lub ankieter;
- błędy proceduralne (kodowania, obliczania itp.)⁶.

Tematyka podjęta w tym artykule odnosi się zatem do błędów systematycznych w badaniach, a uściślając: błędów mierzenia, których źródłem jest przede wszystkim ankieter, a także błędów proceduralnych. Błędy systematyczne, zgodnie z K. Szaniawskim, można ograniczyć między innymi „przez zastosowanie lepszej procedury badawczej: [...] podniesienie jakości pracy terenowej czy też kodowania danych [...]”⁷.

⁴ Nie są to jedyne koncepcje typologii błędów, ich źródeł i relacji między nimi. Więcej informacji na ten temat: P. Jabkowski, *Reprezentatywność badań reprezentatywnych. Analiza wybranych problemów metodologicznych oraz praktycznych w paradygmacie całkowitego błędu pomiaru*, Poznań 2015, s. 21–46, a także F. Sztabiński, *Ocena jakości danych w badaniach surveyowych*, Warszawa 2011, s. 53–60. Celem porównania warto dodać, że na gruncie analiz weryfikacyjnych Jan Lutyński używa takich pojęć, jak: błąd netto, błąd brutto i błąd wyrównany. Są one swego rodzaju odwzorowaniem schematu podanego w publikacji pod redakcją K. Szaniawskiego *Metody statystyczne w socjologii: wybrane zagadnienia*, Warszawa 1968, ale dotyczą tylko weryfikacji zewnętrznej, które to pojęcie jest wytłumaczone w niniejszym tekście. Więcej informacji: J. Lutyński, *Analizy weryfikacyjne w badaniach z zastosowaniem wywiadu kwestionariuszowego, ich rodzaje i możliwości*, [w:] Z. Gostkowski, J. Lutyński (red.), *Studia pilotażowe i analizy weryfikacyjne*, Analizy i Próby Technik Badawczych w Socjologii, t. V, Wrocław 1975, s. 338–339.

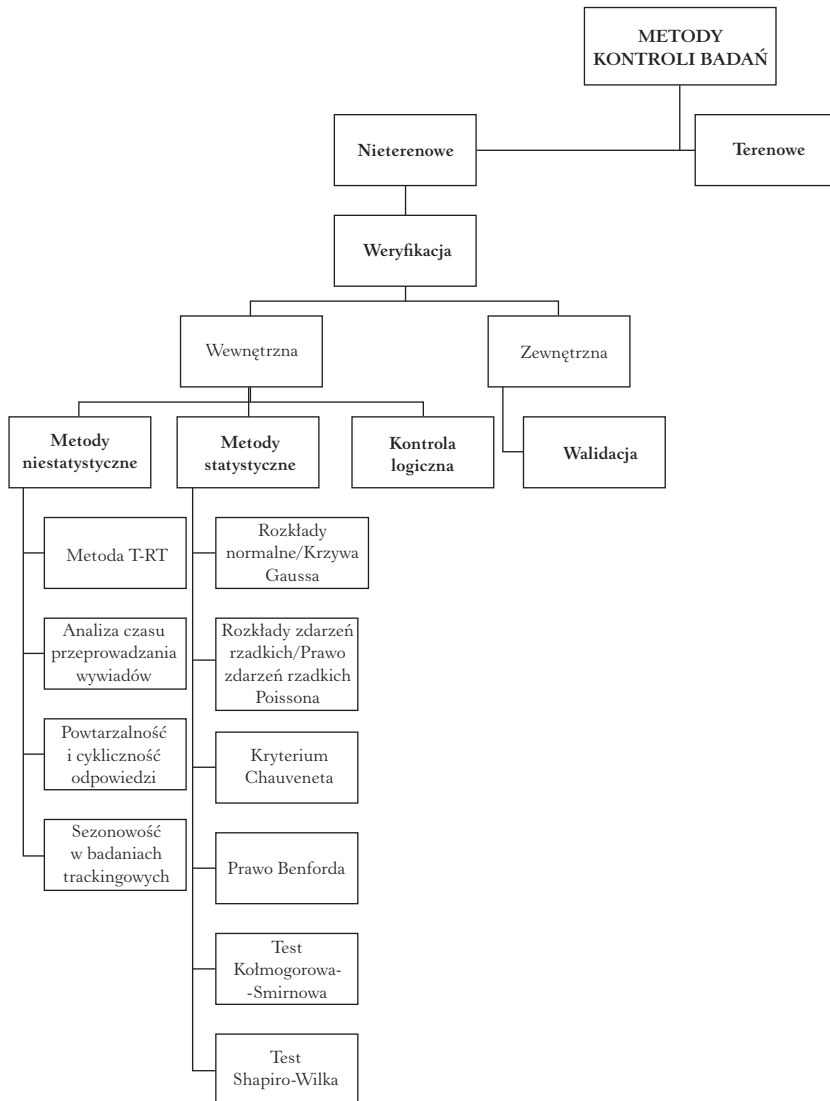
⁵ Ten problem rozwinął w innej pracy pod redakcją K. Szaniawskiego socjolog Grzegorz Lissowski. Przy jego wyjaśnianiu posłużył się także tą samą klasyfikacją błędów co w *Metodach statystycznych w socjologii. Wybrane zagadnienia* (K. Szaniawski (red.), Warszawa 1968). Więcej informacji: G. Lissowski, *Problem jednostek niedostępnych w reprezentacyjnych badaniach socjologicznych*, [w:] K. Szaniawski (red.), *Metody matematyczne w socjologii. Zagadnienia wybrane*, Wrocław 1971, s. 7–34.

⁶ K. Szaniawski (red.), *Metody statystyczne...*, s. 68. Z powyższego wynika, że błędu systematycznego – a co za tym idzie całkowitego błędu badania – nie można w pełni skwantyfikować.

⁷ Tamże, s. 69.

Techniki oceny jakości danych przedstawione w tym artykule służą właśnie tym celom. Zależności między nimi przedstawia rysunek 2.

Rysunek 2. Zależności pomiędzy metodami kontroli badań (ocenami jakości danych)



Źródło: opracowanie własne na podstawie: F. Sztabiński, *Kontrola realizacji badania*, [w:] Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie*, Warszawa 2005, s. 355–358; F. Sztabiński, *Ocena jakości danych w badaniach surveyowych*, Warszawa 2011, s. 97–113; H.M. Blalock, *Statystyka dla socjologów*, Warszawa 1977, s. 92–99, 232–234; M. Rószkiewicz, *Statystyka*, Warszawa 2002, s. 50–74.

Czym różni się walidacja od weryfikacji i kontroli logicznej? Te trzy pojęcia brzmią podobnie i osobie niewtajemniczonej w temat analizy wiarygodności danych może wydawać się, że mogą być stosowane wymiennie. Tak jednak nie jest, a wyjaśnienie niuansów między analizowanymi pojęciami będzie przedmiotem tej części artykułu. Istotne jest przy tym podkreślenie, że kontekstem omawianych treści, zarówno w tej części jak i następczej, będą wyłącznie dane ilościowe pozyskiwane w badaniach socjologicznych z użyciem wywiadów kwestionariuszowych⁸.

Wszelkie metody kontroli wiarygodności danych opisane w części pierwszej i drugiej artykułu należą do tzw. metod nieterenowych. Ale metody statystyczne i stochastyczne⁹ także zaliczają się do tej grupy. Kontrola nieterenowa¹⁰ – w przeciwieństwie do terenowej – opiera się na

⁸ Po 1989 roku, kiedy do Polski zaczęły docierać nowoczesne technologie służące obróbce danych ilościowych (sprzęt, ale przede wszystkim oprogramowanie różnego rodzaju), łatwiejsze stały się metody walidowania, weryfikowania i logicznej kontroli danych, ale przede wszystkim – bardziej precyzyjne. Można byłoby zatem założyć, że w polskich naukach społecznych (jak też na rynku badań ilościowych i jakościowych) dopiero od lat 90. XX w. poruszany jest problem walidacji, weryfikacji i kontroli logicznej. Jest to założenie błędne, gdyż pierwsze znaczące teksty i badania w tym zakresie powstawały już pod koniec lat 60. XX w. Szczególny wkład ma tutaj tzw. łódzka szkoła metodologiczna na czele z badaczami takimi jak Jan Lutyński, a w późniejszym okresie także Franciszek Sztabiński. Nie są to oczywiście wszyscy socjologowie z tej grupy – tutaj należy wymienić również Pawła Daniłowicza, Zygmunta Gostkowskiego, Jerzego Koniarka, Krystynę Lutyńską czy Pawła Sztabińskiego – natomiast to teksty przede wszystkim wcześniej wymienionych autorów są podstawą tej części artykułu. Więcej o łódzkiej szkole metodologicznej na stronie Instytutu Socjologii UŁ: Katedra Metod i Technik Badań Społecznych, *O katedrze*, <http://instytutsocjologii.uni.lodz.pl/instytut/katedry-i-zaklady/katedra-metod-technik-badan-spolecznych/> [dostęp: 10.01.2019]. Z kolei najważniejsze teksty poruszające temat artykułu znajdują się w serii *Analizy i Próby Technik Badawczych w Socjologii* wydawanej przez Polską Akademię Nauk, Instytut Filozofii i Socjologii oraz Uniwersytet Łódzki, Instytut Socjologii, warto tu wspomnieć tom 13: *Analizy weryfikacyjne – przeszłe i obecne doświadczenia badawcze* (Łódź 2017) pod redakcją Katarzyny Grzeszkiewicz-Radulskiej i Anety Krzewińskiej; także w czasopiśmie „ASK. Research and Methods” (IFiS PAN); ponadto w publikacjach: Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork jest sztuką*, Warszawa 2005 oraz przywołanej już pracy: F. Sztabiński, *Ocena jakości danych w badaniach surveyowych*.

⁹ Metody te, co więcej, zaliczają się do tzw. weryfikacji wewnętrznej, o czym będzie dalej mowa.

¹⁰ Warto dodać, że metody kontroli służą także dwu celom: po pierwsze, poprawieniu danych lub ocenie jakości danych znajdujących się w bazie, ale także, po drugie, ocenie ankietowanych uczestniczących w procesie pozyskiwania informacji. Artykuł ten skupia się jednak tylko na tym pierwszym kontekście, aczkolwiek warto podkreślić, że oba te cele się wzajemnie przenikają. Więcej informacji dotyczących drugiego celu można zna-

takiej analizie danych, która nie wymaga ponownego kontaktu z respondentem celem ustalenia, jakiej udzielił odpowiedzi. Ze względu na zakres omawianych pojęć weryfikacja jest zdecydowanie terminem najszerszym, dlatego od niej rozpoczniemy. Co więcej, walidację można uznać za specyficzną metodę tzw. weryfikacji zewnętrznej, a kontrolę logiczną – weryfikacji wewnętrznej.

Weryfikacja

Analizy weryfikacyjne – zgodnie z Janem Lutyńskim – można podzielić na dwa główne typy: wewnętrzną i zewnętrzną¹¹. Typ pierwszy dostarcza wiedzy o procesach otrzymywania informacji za pomocą narzędzia badawczego. Innymi słowy w pierwszej kolejności ocenie podlega proces, a dopiero potem oceniane są same wyniki badania¹². W przypadku drugiego typu weryfikacji kolejność jest odwrotna, to jest na podstawie danych uzyskanych inną techniką niż założona w badaniu (lub podczas innego pomiaru wykonywanego tą samą techniką) można ocenić, czy wybrana technika (pomiar) przyniosła wiarygodne wyniki. Warto przy tym nadmienić, że analizy weryfikacyjne oprócz informacji o metodologii mogą także pomóc przy ocenie wielkości błędów popełnionych w procesie pozyskiwania danych, a nawet poprawieniu tych błędów, to jest tzw. błędnych zaklasyfikowań¹³.

Weryfikacja wewnętrzna jest metodą, którą trudniej stosować w ocenie jakości danych. Ponadto nie zawsze musi przynosić wyniki liczbowe (rozumiane jako liczba bądź odsetek błędnych zaklasyfikowań), a co za tym idzie – prowadzić do poprawek w zestawieniu danych. Nie oznacza to, że przy tym typie weryfikacji niemożliwe jest używanie technik

leżać szczególnie w publikacji: Z. Sawiński, F. Sztabiński, *Czy ankieterzy oszukują? Jak można to sprawdzić?*, [w:] Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork...*, s. 361–380.

¹¹ J. Lutyński, *Analizy weryfikacyjne...*, s. 330. Należy dodać, że w zupełnie innym, znacznie szerszym kontekście użył tego terminu Stefan Nowak w swojej publikacji *Metody badań socjologicznych*. Zgodnie z nim celem badań weryfikacyjnych jest „sprawdzenie empiryczne, kontrola prawdziwości jednego bądź też całego zespołu ogólnych twierdzeń o związkach między pewnymi, ogólnie zdefiniowanymi klasami zjawisk. W badaniu takim rzeczywistość przebadana interesuje nas jedynie jako próba tej ogólnie zdefiniowanej klasy przedmiotów, do których odnieść się ma nasze twierdzenie”. Więcej informacji: S. Nowak, *Metody badań socjologicznych*, Warszawa 1965, s. 191–193.

¹² Tamże, s. 362; por. F. Sztabiński, *Ocena jakości...*, s. 79.

¹³ J. Lutyński, *Analizy weryfikacyjne...*, s. 340.

matematycznych, ponieważ te analizy weryfikacyjne dzielą się na dwa główne typy: niestatystyczne i statystyczne. O tych drugich będzie mowa w dalszych częściach niniejszego artykułu.

„Niestatystyczne metody weryfikacji wewnętrznej polegają [...] na szacowaniu, na ile coś jest prawdopodobne”¹⁴. Pomocne przy takich weryfikacjach może być tworzenie tzw. normatywnych modeli formowania informacji¹⁵. Modele budować należy według tego, co J. Lutyński nazwał metodą „logiczną”¹⁶. Polega ona na określeniu przesłanek służących przy porządkowaniu danej jednostki badania do arbitralnie stworzonych przez badacza (lub analityka) klas respondentów. Rysunek 3 przedstawia przykład zastosowania takiego uproszczonego modelu (pogrubieniem zaznaczono odpowiedzi respondenta).

Jak wykazuje rysunek 3, respondent udzielił odpowiedzi nietypowej, to jest niezgodnej z przewidywaniami badacza. Co bardzo ważne – jeśli założenie badacza nie zostało spełnione, nie oznacza to od razu, że w zbiorze danych pojawił się błąd. Oznacza to jedynie, że – nadal – korzystając z technik nieterenowych można sprawdzić, czy respondent ze względu na swoje cechy, najczęściej społeczno-demograficzne, udzielił odpowiedzi typowej dla respondentów zbliżonych parametrami. Metoda ta określana jest jako model najbliższego sąsiedztwa¹⁷. Jeśli natomiast wyniki takiej analizy wskażą, że odpowiedzi respondenta stanowią pewną anomalię to jest to wskazaniem do sprawdzenia reszty kwestionariuszy dostarczonych przez ankietera. W ostateczności konieczny może być ponowny kontakt z respondentem (telefoniczny lub inny, to jest wymagający skorzystania z terenowej metody kontroli¹⁸).

¹⁴ W tym wypadku nie ma oczywiście mowy o prawdopodobieństwie matematycznym, a jedynie jego intuicyjnym rozumieniu. Por.: F. Sztabiński, *Ocena jakości...*, s. 79. Wynika z tego wniosek, że ten typ weryfikacji może być – lub wręcz powinien być – stosowany także przed fazą realizacji badania.

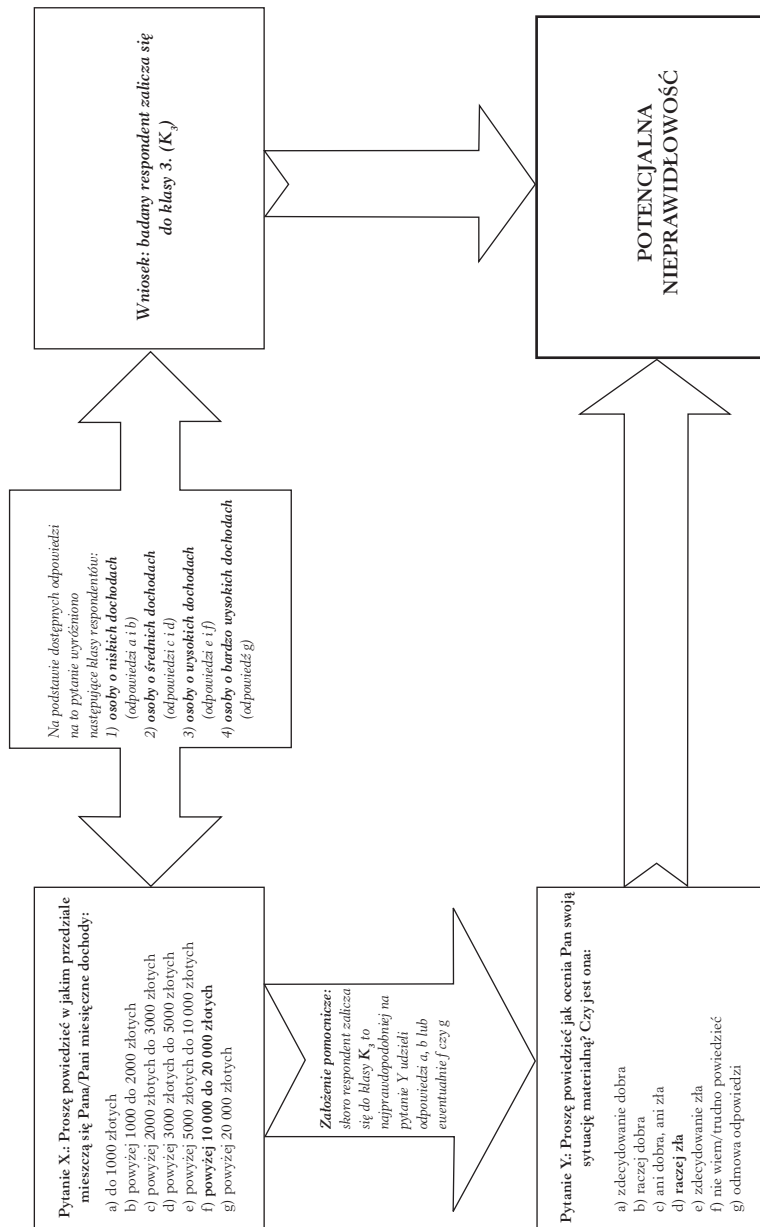
¹⁵ J. Lutyński, *Analizy weryfikacyjne...*, s. 367–370.

¹⁶ Tamże, s. 370–380.

¹⁷ Model najbliższego sąsiedztwa jest modelem ekonometrycznym, który został stworzony przez Konrada Kuźmę w 2015 roku na podstawie – służącego do rozwiązywania problemu komiwojażera – algorytmu najbliższego sąsiedztwa. Model opiera się na analizie regresji. Więcej na ten temat można przeczytać np. tutaj: *StatSoft (2006). Elektroniczny Podręcznik Statystyki PL*, Kraków, https://www.statsoft.pl/textbook/stathome_stat.html (dostęp: 10.01.2019).

¹⁸ Informacje na temat terenowych metod kontroli można znaleźć m.in. w: F. Sztabiński, *Kontrola realizacji badania*, [w:] Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fiel-dwork...*, s. 356–358 oraz: tenże, *Ocena jakości...*, s. 105–113.

Rysunek 3. Przykład normatywnego modelu formowania informacji
(Przeгляд schematu należy rozpocząć od prostokąta z pytaniem X. Pogrubieniem zaznaczono odpowiedzi respondenta)



Źródło: opracowanie własne.

Weryfikacja zewnętrzna zakłada porównanie zbiorów danych, które różnią się od siebie znanym wcześniej (lub domniemanym) stopniem pewności (wiarygodności). Można pod tym względem wyróżnić trzy typy takich analiz¹⁹:

- wyniki ze źródła, do którego porównujemy inne rezultaty, są praktycznie pewne (przykładem może być porównanie deklarowanego wieku respondentów z ich danymi zapisanymi w bazie PESEL, a uzyskanymi podczas losowania próby na podstawie tego operatu);
- wyniki ze źródła, do którego porównujemy, są bardziej pewne niż wyniki z wywiadu²⁰ (taka analiza mogłaby się opierać na porównaniu deklaracji pracowników przedsiębiorstwa na temat spóźnień i wywiadów z ich kierownikami (bez użycia tzw. danych dokumentalnych²¹, czyli w tym wypadku na przykład rejestru godzin wejść i wyjść);
- nie jest wiadome, jak pewne są rezultaty z obu źródeł²² (ta analiza z kolei mogłaby dotyczyć preferencji partyjnych wyborców, przy czym jedno badanie byłoby oparte na próbie kwotowej, a inne na losowej).

Analizy weryfikacyjne można podzielić także w sposób przedstawiony w tabeli 1. W tym miejscu warto objaśnić pojęcia składowe w niej użyte. Metoda nie jest w tym wypadku równoznaczna z pomiarem, ponieważ analizy weryfikacyjne pierwszego i drugiego typu można wykonywać z użyciem tego samego narzędzia badawczego (por. z dalej opisaną metodą T-RT). Inaczej jest w przypadku typu trzeciego, którego istotą jest obliczanie korelacji²³ między wynikami z różnych źródeł²⁴. Ważne jest

¹⁹ Opracowanie własne na podstawie: J. Lutyński, *Analizy weryfikacyjne...*, s. 337.

²⁰ Konieczne jest w tym przypadku, jak i poprzednim, uzasadnienie, dlaczego jedno ze źródeł jest bardziej pewne od drugiego.

²¹ Więcej na ten temat czytaj np. w: F. Sztabiński, *Ocena jakości...*, s. 89.

²² Więcej informacji na temat porównywania wyników niepewnych: J. Lutyński, *Analizy weryfikacyjne...*, s. 346–358.

²³ Pomimo że J. Lutyński proponuje w swoim rozwiązaniu obliczanie korelacji, nie jest to jedyna metoda. Możliwym rozwiązaniem byłaby również regresja lub analiza wariancji z powtórzeniami. Więcej na ten temat można znaleźć np. w: *StatSoft (2006). Elektryczny Podręcznik...*

²⁴ W tym wypadku dobrym przykładem są tzw. sondaże *exit poll*, czyli badanie kwestionariuszowe przeprowadzane z losowo dobranymi wyborcami w dniu wyborów. Badanie takie nie dotyczy deklaracji, jak respondent zamierza zagłosować (preferencji wyborczych), lecz tego, jak już zagłosował. Jego celem jest oszacowanie wyników głosowania i wyniki takiego badania są z reguły upubliczniane (np. przez stacje telewizyjne) tuż po zamknięciu lokali wyborczych, kiedy jeszcze nie są możliwe do obliczenia i rozpowszechnienia rzeczywiste rezultaty elekcji. Więcej informacji: R. Hilmer, *Exit Polls – A Lot More than Just a Tool for Election Forecasts*, [w:] M. Carballo, U. Hjelm (red.), *Public Opinion Polling in a Globalized World*, Berlin 2008, s. 95. Jeśli wyniki takiego sondażu pokrywają

Tabela 1. Podział weryfikacyjnych analiz zewnętrznych bez uwzględnienia kryterium pewności źródeł

Numer typu	Metody ^a	Powtórzony pomiar z użyciem tego samego narzędzia	Zbiorowość	Możliwość bezpośredniej identyfikacji respondentów	Próba	Dopuszczalne wnioski
1a	Mogą być różne	Możliwy	Ta sama	Wymagana	Musi być taka sama	Liczba błędnych zaklasyfikowań z możliwością ich identyfikacji
1b				Wykluczona	Ten sam typ (np. losowa), ale różne jednostki	Liczba błędnych zaklasyfikowań
2				Opcjonalna	Może być różna	Tylko dotyczące pewności obu źródeł danych (jedno z nich jest pewniejsze, a drugie mniej pewne)
3	Muszą być różne	Niemożliwy				

^a Opracowanie dotyczy tylko danych ilościowych (zgodnie z kontekstem opisanym w: J. Lutyński, *Analizy weryfikacyjne...*, s. 319–325), ale należy mieć na uwadze, że dane weryfikuje się także w badaniach jakościowych.

Źródło: opracowanie własne na podstawie J. Lutyński, *Analizy weryfikacyjne...*, s. 333–337.

tu ukryte założenie, że korelacja między różnymi pomiarami z użyciem tego samego narzędzia byłaby bardzo wysoka, o ile nie równa jeden²⁵.

W przypadku analiz typu pierwszego chodzi o co najmniej dwukrotny kontakt z dokładnie takimi samymi jednostkami badania. Interesujący jest tutaj problem identyfikacji respondentów, ponieważ przy założeniu, że ich identyfikacja jest zapewniona (1a), możliwa jest dokładna weryfikacja danych co do każdego przypadku (respondenta)²⁶. Przy braku identyfikacji (1b) wiadomo jedynie, na ile różnią się pierwotne wyniki od powtórzonych, ale nie jest możliwe określenie, których badanych należy zaklasyfikować inaczej²⁷. Natomiast w przypadku analizy drugiego typu jej istotą jest porównanie wyników uzyskanych z całkowicie różnych prób, dlatego, nawet jeśli możliwa jest identyfikacja respondentów, to wykluczony jest ponowny wywiad z nimi²⁸.

Opisywane trzy typy analiz dają pole do różnych wniosków. W najwyższym stopniu mogą podnieść jakość danych analizy typu 1a, ponieważ oprócz określenia skali rozbieżności (w liczbach czy procentach) możliwe jest – dzięki identyfikacji respondentów – poprawienie błędnych zaklasyfikowań z pierwszego pomiaru. Natomiast analizy typu 1b i 2 dostarczają tylko informacji, jaka jest wielkość niewłaściwych wskazań, ale nie będzie uprawnione na ich podstawie korygowanie danych w zbiorze. Najmniej użyteczne pod tym względem są analizy korelacyjne (typ 3), ponieważ na ich podstawie można jedynie ocenić, która z przyjętych

się w dużej mierze (czyli korelacja jest wysoka) z faktycznymi rezultatami wyborów, oznacza to, że rezultaty badania są wiarygodne. Jak pokazują sondaże *exit poll* prowadzone przez zespół Pawła Predki w TNS OBOP, a obecnie IPSOS Polska, wyniki są, szczególnie w przypadku wyborów prezydenckich, prawie identyczne jak wyniki PKW. Najbardziej widoczne było to w czasie przyspieszonych wyborów prezydenckich 2010 r., kiedy to różnica – w II turze – pomiędzy wynikami *exit poll* a PKW wynosiła 0,01%.

²⁵ Nie znaczy to jednak, że tak jest zawsze, o czym mowa w dalszej części tekstu. Co więcej, dokładnie to samo założenie stoi za analizami weryfikacyjnymi typu 1. i 2. (oczywiście tylko w wypadku użycia identycznego narzędzia).

²⁶ Choć jak pokazuje praktyka badawcza, nie zawsze tak się dzieje.

²⁷ Przykładem analizy opartej na próbie umożliwiającej identyfikację respondentów jest np. zapytanie o znajomość pisarza, a następnie o wymienienie kilku jego publikacji. W drugim przypadku można wyobrazić sobie wywiad kwestionariuszowy z pytaniem o znajomość pisarza, a następnie (kolejny pomiar) – podczas w pełni anonimowej ankiety audytoryjnej – pytanie o wymienienie kilku dzieł. W obu przypadkach podstawą badania są ci sami respondenci.

²⁸ Przykładem takiej analizy jest wykonanie w tym samym czasie badania dotyczącego znajomości marek produktów, opartego na dwóch różnych próbach losowych (to jest dwóch wylosowanych w ten sam sposób próbach, przy tym żaden z respondentów nie znajduje się w obu próbach jednocześnie).

metod jest mniej (lub bardziej) pewna oraz jaka jest skala i kierunek zmian w zbiorze danych w porównaniu jednej metody z drugą²⁹.

Na koniec należy powiedzieć, że wyniki z weryfikacji wewnętrznej i zewnętrznej nie muszą się ze sobą pokrywać³⁰. Jeśli tak nie jest, oznacza to, że rezultaty badania należy dokładniej sprawdzić, używając innych sposobów.

Walidacja

Walidacja, jak wskazuje Franciszek Sztabiński za Hughiem Parrym i Helen Crossley, może być uznawana za formę weryfikacji wyników badania. W tym wypadku byłaby to weryfikacja dokumentalna³¹, czyli porównanie danych faktycznych z oficjalnymi rejestrami urzędowymi uznawanymi za całkowicie pewne źródło informacji³². Innymi słowy, walidacja jest specjalnym typem weryfikacji zewnętrznej. W wymienionej definicji zawarte jest także istotne ograniczenie walidacji: tylko niektóre typy danych mogą podlegać tej metodzie³³. Przykładowo, do pytań dotyczących postaw czy opinii nie będzie można jej zastosować.

Warto podkreślić, że w innej swojej publikacji F. Sztabiński definiuje weryfikację bez rozróżniania na podtypy, w sposób tożsamy z walidacją³⁴. Ponadto w tej samej publikacji wyróżnia funkcję walidacyjną³⁵ jako jedną z funkcji kontroli realizacji badania. Jest to „ocena badania w kategoriach rzetelności i wiarygodności jego wyników”³⁶.

²⁹ F. Sztabiński, *Ocena jakości...*, s. 83–84.

³⁰ Dobrym zilustrowaniem takiej sytuacji jest dowolne badanie przeprowadzane na studentach, w którym muszą oni podać swój rok urodzenia. Można założyć, że jeśli ktoś jest studentem, to nie może być starszy niż x lat, ale nie może być młodszy niż y lat ($x > y$; weryfikacja wewnętrzna). Jeśli w zbiorze danych pojawi się nietypowy rezultat, to sięgając do oficjalnych rejestrów udostępnionych przez uczelnię można dokonać weryfikacji zewnętrznej i pozostawić albo usunąć daną obserwację z bazy. Jeśli natomiast konstrukcja próby umożliwiałaby ponowny bezpośredni kontakt z takim respondentem, to sięgając do terenowych metod kontroli, można byłoby także w ten sposób zweryfikować ów wynik. Więcej informacji: J. Lutyński, *Analizy weryfikacyjne...*, s. 387–392.

³¹ F. Sztabiński, *Ocena jakości...*, s. 78; por. tamże, s. 85.

³² Tamże, s. 78.

³³ Tamże, s. 85–86.

³⁴ F. Sztabiński, *Kontrola realizacji...*, s. 356.

³⁵ Tamże, s. 350–351.

³⁶ Tamże, s. 350. Rzetelność może być uznana za cechę jakości pomiaru i „mówi o tym, z jaką dokładnością zmierzylimy to, co zmierzylimy, lub dokładność z jaką test (bada-

Ten sam autor wskazuje, że istnieje także możliwość walidowania danych na podstawie rzeczywistego zachowania badanych³⁷. Niemniej, celem uporządkowania pojęć, w niniejszym artykule przyjęto definicję walidacji zaproponowaną przez H. Parry'ego i H. Crossley. Schematyczny, uproszczony przykład procedury walidowania danych pokazują tabele 2 i 3.

Tabela 2. Pierwszy etap walidacji danych – przykładowe pytanie kontrolne

<p>Pytanie Z: Przed nami już ostatnie pytanie. Jest to pytanie służące wyłącznie kontroli mojej pracy przez przełożonych, dlatego byłbym wdzięczny za udzielenie odpowiedzi. Proszę o podanie swojej dokładnej daty urodzenia i miejscowości, w której się Pan/Pani urodził(-a):</p>	
Dzień (DD):	13
Miesiąc (MM):	07
Rok (RRRR):	87
Nazwa miejscowości urodzenia:	gdansk

Źródło: opracowanie własne.

Tabela 3. Drugi etap walidacji danych – porównanie danych wpisanych przez ankietera z danymi z operatu losowania

Rodzaj danych	Metryczka respondenta nr xxx (z próby)	Metryczka kontrolna wpisana przez ankietera
Dzień (DD):	13	13
Miesiąc (MM):	08	07
Rok (RRRR):	1987	87
Nazwa miejscowości urodzenia:	Gdańsk	gdansk

Źródło: opracowanie własne.

nie, narzędzie) mierzy to, co mierzy”; F. Sztabiński, *Ocena jakości...*, s. 71. Z kolei z pojęciem rzetelności jest mocno związane pojęcie trafności, które także może być kryterium jakości pomiaru: „mówi o tym, czy narzędzie mierzy to, co ma mierzyć lub z jaką dokładnością zmierzylimy to, co chcieliśmy zmierzyć”; tamże, s. 62. Wynika z tego, że trafność jest pojęciem szerszym niż rzetelność. Więcej informacji: tamże, s. 61–76.

³⁷ Tamże, s. 89–90.

Tabela 4. Trzeci etap walidacji danych. Poprawienie danych wpisanych przez ankietera

ID respondenta	Dzień	Miesiąc	Rok	Nazwa miejscowości urodzenia
XXX	13	08	1987	Gdańsk

Źródło: opracowanie własne.

Jak widać na podanym przykładzie, ankieter, prócz pomijalnych błędów, dobrze wpisał dane respondenta, co każe sądzić, że – przynajmniej w rozmowie z tym badanym – rzetelnie wykonał swoją pracę.

Kontrola logiczna

Kontrola logiczna jest specyficznym typem weryfikacji wewnętrznej³⁸ i odnosi się do merytorycznej analizy wypełnionych kwestionariuszy³⁹. Dalsza część opracowana została na podstawie materiałów wewnętrznych firm CBM INDICATOR oraz TNS OBOP. Kontrola ta dotyczy dwóch aspektów odpowiedzi respondentów, które – na podstawie naszego doświadczenia – nazwaliśmy⁴⁰:

- aspektem formalnym: sprawdzanie wewnętrznej logicznej spójności odpowiedzi (na przykład ich zgodności z filtrami), jak również na przykład ustalenie, czy ankieter zadał wszystkie pytania zawarte w kwestionariuszu;
- aspektem nieformalnym: weryfikacja „zależności między odpowiedziami na różne pytania, które w mniejszym lub większym stopniu związane są z kryteriami losowania lub doboru”⁴¹. W tym przypadku ten proces jest niemal tożsamy z normatywnymi modelami formowania informacji.

Warto podkreślić, że pilotaż badania odpowiednio przeprowadzony przed fazą realizacji pomaga uniknąć lub znacznie ograniczyć później-

³⁸ Świadczy o tym przede wszystkim fakt, że nie korzysta się podczas niej z żadnych danych zewnętrznych.

³⁹ F. Sztabiński, *Ocena jakości...*, s. 101.

⁴⁰ Autorem nazw jest Konrad Gałuszko, pracownik firmy CBM Indicator i współautor artykułu. Nazwy zostały nadane, ponieważ w materiałach CBM Indicator nie określono ich dla tych dwóch aspektów kontroli logicznej.

⁴¹ F. Sztabiński, *Ocena jakości...*, s.101.

sze błędy w zbiorze danych⁴². Należy zatem uznać, że pilotaż przed badaniem będzie dotyczyć przede wszystkim samego narzędzia badawczego, a kontrola logiczna po zebraniu danych – ocenie jakości danych czy ankierów. Inaczej mówiąc, kontrola logiczna jest przydatna do wskazania, które podmioty badania – badacze, ankierzy czy respondenci – są źródłem błędów. Jeśli badacze nie zauważyli błędów w kwestionariuszu podczas jego wieloetapowej analizy oraz pilotażu, oznacza to, że za późniejsze niezgodności w zbiorze danych są odpowiedzialni przede wszystkim badacze. W przypadku poprawnej konstrukcji, jeśli stwierdzono błędy (już po zakończeniu realizacji badania), oznacza to, że najprawdopodobniej ankierzy – celowo lub przypadkowo – dokonali pomyłki. Ponadto, zdaniem niektórych badaczy, to na ankierze spoczywa obowiązek stałego nadzorowania przebiegu wywiadu i sprawdzania, czy sam respondent udziela spójnych logicznie, najbardziej prawdopodobnych odpowiedzi. Aczkolwiek na przestrzeni lat założenie to uległo zmianie i obecnie w większości prowadzonych projektów odpowiedzi respondentów pozostawia się nienaruszone. Jeśli zatem kwestionariusz jest poprawny, a ankier wiarygodny, oznacza to, że źródłem błędu jest najprawdopodobniej sam respondent⁴³.

Obecnie dane z badań są przechowywane głównie w formie elektronicznych baz danych, dlatego należy dodać, że kontrola logiczna jest związana z tzw. redundancją danych. Pojęcie to pochodzi z zakresu informatyki i oznacza „nadmiarowość; praktykę przechowywania więcej niż jednego elementu tych samych danych (w kilku kopiach)”⁴⁴. Aby ograniczyć redundancję, stosuje się tzw. *data reduction plan*, czyli „plan redukcji danych; zestaw instrukcji dla edycji i kodowania kwestionariuszy i ankier oraz analizy i weryfikacji udzielanych odpowiedzi i uzyskiwanych danych”⁴⁵. Kontrola logiczna, jako działanie redukujące redundancję, pozwala

⁴² Oczywiście obecnie programy i komputery uniemożliwiają w dużym stopniu popełnienie błędów formalnych – zarówno przed, jak i po realizacji badania. Należy zatem przyjąć, że w niniejszym tekście rozważania dotyczące formalnej kontroli logicznej dotyczą szczególnie badań PAPI.

⁴³ T. Pawłowski, *Logiczne podstawy weryfikacji wewnętrznej badań kwestionariuszowych*, [w:] Z. Gostkowski, J. Lutyński (red.), *Wywiad kwestionariuszowy w świetle badań metodologicznych*, Analizy i Próby Technik Badawczych w Socjologii, t. 4, Wrocław 1972, s. 277–278. Więcej informacji na ten temat także w: C. Blattman i in., *Measuring the Measurement Error: A Method to Qualitatively Validate Survey Data*, <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31847/Measuring%20the%20measurement%20error.pdf?sequence=2&isAllowed=y> (dostęp: 20.02.2019).

⁴⁴ M. Trojański, *Dictionary of Applied Informatics*, Warszawa 2007, s. 372.

⁴⁵ Tamże, s. 149.

po zakończeniu badania uzupełnić pewne braki lub usunąć nieuzasadnione nadmiary danych. Odwołując się do wcześniejszego przykładu z rysunku 1, badacze podczas weryfikacji wewnętrznej przed badaniem mogą uznać prawdopodobieństwo takiego układu odpowiedzi za niskie. Niemniej respondent miał prawo odpowiedzieć w taki sposób, tak więc decyzja o usunięciu takiej odpowiedzi lub całej obserwacji będzie efektem kontroli logicznej drugiego typu. Natomiast jeśli ankietowany odpowiedział na pytanie kwestionariusza, w sytuacji gdy jeden z filtrów nakazywał ominięcie następnego pytania, a mimo to respondent zaznaczył jakąś odpowiedź (nadmiar danych), jej usunięcie nie wpłynie na późniejsze rezultaty, a nawet poprawi ich wiarygodność (por. z częścią o powtarzalności i cykliczności odpowiedzi). Można wyobrazić sobie też sytuację, w której respondent w jednym miejscu kwestionariusza podaje swoją pełną datę urodzenia, a w innym miejscu pomija pytanie kontrolne o wiek w liczbach. Podczas kontroli danych można taki brak danych uzupełnić⁴⁶.

Metody niestatystyczne

Jak pisaliśmy wcześniej w części poświęconej weryfikacji, metody niestatystyczne bazują na intuicyjnej ocenie prawdopodobieństwa wystąpienia danego układu czy rozkładu odpowiedzi w całym badaniu i/lub w kwestionariuszach dostarczonych przez określonego ankietera⁴⁷. W niniejszej części artykułu będą opisane następujące metody: T-RT, analiza czasu przeprowadzania wywiadów, powtarzalności i cykliczności odpowiedzi, a także sezonowości w badaniach powtarzalnych i ciągłych (w praktyce badawczej nazywanych trackingowymi). Należy mieć na uwadze, że metody statystyczne i niestatystyczne niekiedy się ze sobą

⁴⁶ Przy kontroli logicznej pomocne jest – oprócz normatywnych modeli formowania informacji – zdefiniowanie związków logicznych między poszczególnymi pytaniami i ich kafeteriami. T. Pawłowski wyróżnia następujące związki: 1. czyste: wynikanie logiczne, równoważność logiczna, sprzeczność, przeciwieństwo, podprzeciwieństwo; 2. eliptyczne: wynikanie eliptyczne, równoważność eliptyczna, sprzeczność eliptyczna. Niestety ze względu na ograniczoną objętość tekstu nie jest możliwe dokładne ich omówienie. Należy dodać, że przy określaniu związków logicznych pomocne mogą być także tabele zestawiające relacje między poszczególnymi pytaniami. Więcej informacji: T. Pawłowski, *Logiczne podstawy weryfikacji wewnętrznej badań kwestionariuszowych*, [w:] Z. Gostkowski, J. Lutyński (red.), *Wywiad kwestionariuszowy w świetle badań metodologicznych*, Analizy i Próby Technik Badawczych w Socjologii, t. 4, Wrocław 1972, s. 279–290.

⁴⁷ F. Sztabiński, *Ocena jakości...*, s.79; Z. Sawiński, F. Sztabiński, *Czy ankieterzy...*, s. 361–380.

zazębiają – na przykład aby wydać werdykt dotyczący jakości danych za pomocą techniki niestatystycznej, należy zastosować metodę statystyczną, ponieważ czasem zwykły „rzut oka” nie będzie wystarczający. W psychologii – skąd pochodzi to pojęcie – jest to działanie nazywane kwantyfikacją wyników. Dobrym przykładem mogą tu być analizy powtarzalności i cykliczności odpowiedzi, a także sezonowości w badaniach powtarzalnych i ciągłych.

METODA T-RT

Nazwa metody pochodzi od angielskiego skrótu oznaczającego *test-retest*⁴⁸. Wymiennym określeniem jest „powtórzony pomiar”. T-RT jest wskaźnikiem oceniającym między innymi rzetelność pomiaru⁴⁹, ale także narzędzia, za pomocą którego go dokonano, lub elementu tego narzędzia (na przykład skali)⁵⁰. Korzenie tej metody można odnaleźć w naukach przyrodniczych, w których powtarzalność rezultatów jest jedną z wymaganych cech pomiaru lub narzędzia⁵¹.

Istotą metody powtórzonego pomiaru jest stabilność wyników w czasie⁵². Należy podkreślić przy tym, że oba pomiary muszą być dokonane przy użyciu identycznego narzędzia badawczego. Jeśli wyniki uzyskane za jego pomocą są takie same lub zbliżone, oznacza to, że pomiar i/lub narzędzie są wiarygodne⁵³.

Metoda ta może jednak przynosić błędne, nieuzasadnione konkluzje, ponieważ jest wrażliwa na trzy następujące czynniki: wpływ czasu między pomiarami, warunki zewnętrzne wpływające na respondentów⁵⁴, a także

⁴⁸ Takie rozwinięcie tego skrótu pada m.in. w: H. Domański, A. Dukaczewska, *Stabilność odpowiedzi w badaniach socjologicznych*, „ASK. Research and Methods” 1996, nr 1, s. 74, a także w: F. Sztabiński, *Ocena jakości...*, s. 72.

⁴⁹ F. Sztabiński, *Ocena jakości...*, s. 72.

⁵⁰ H. Domański, A. Dukaczewska, *Stabilność odpowiedzi...*, s. 71.

⁵¹ F. Sztabiński, *Ocena jakości...*, s. 72.

⁵² W przypadku małych prób i/lub zjawisk o niskim rozpowszechnieniu stabilność ta może być zaburzona, dlatego zbadanie np. długofalowych trendów może przynieść błędne wnioski. Aby tego uniknąć, należy – oprócz zachowania odpowiednio długiego okresu kontynuacji badania (np. trackingu) – dokonać zabiegu tzw. agregacji podłużnej, zwanego też rolowaniem danych. Polega on na łączeniu ze sobą danych pochodzących z kilku kolejnych transz (punktów pomiarowych). Z. Sawiński, *Badania trackingowe*, [w:] D. Maison, A. Noga-Bogomilski, *Badania marketingowe. Od teorii do praktyki*, Gdańsk 2007, s. 112–114.

⁵³ H. Domański, A. Dukaczewska, *Stabilność odpowiedzi...*, s. 71.

⁵⁴ Dobrym przykładem w badaniach opinii może tu być tzw. zjawisko *agenda-setting*. Autorami tego pojęcia, jak i pierwszych studiów na ten temat są Marshall McCombs

tzw. zjawisko *pre-testu* i *post-testu*⁵⁵. Z tej też przyczyny T-RT może być zastosowany tylko przy zachowaniu warunków zdefiniowanych przez Marka Stycznia, które w swojej pracy przytacza F. Sztabiński:

- „oba pomiary winny być dokonane w takim odstępie czasu, aby mierzone własności nie mogły ulec zmianie;
- dystans czasowy między tymi dwoma pomiarami winien być na tyle długi, aby na pomiar nie wpływały te same, niekontrolowane czynniki;

i Donald Shaw. Zgodnie z hipotezą *agenda-setting*, „media mają zdolność narzucania opinii publicznej przekonania co do ważności określonych tematów – kwestii, problemów, a to sprawia, że obywatele uważają za ważne te same tematy, o których najczęściej mówi się w mediach”; E. Nowak, *Teoria „agenda setting” a nowe media*, „Studia Medioznawcze” 2016, nr 3(66), s. 13.

⁵⁵ H. Domański, A. Dukaczewska, *Stabilność odpowiedzi...*, s. 73–74. Zjawisko to oznacza wpływ samego badania, a zwłaszcza pierwszego pomiaru (pretestu) – a nie czynników zewnętrznych czy innych – na uzyskiwane odpowiedzi w dalszych pomiarach (posttestach). Innymi słowy na stabilność lub zmienność wyników w czasie może wpływać podczas procedury T-RT to, że respondenci „aktywizują się poznawczo: zwracają uwagę na pewne sprawy, skłaniają się do szukania nowych o nich [przedmiocie badania] informacji [...]”; A. Sulek, *Eksperyment w badaniach społecznych*, Warszawa 1979, s. 76. Dobrym przykładem ilustrującym wpływ pretestu jest tzw. efekt panelowy albo „uczenie się” odpowiedzi. Pierwsze zjawisko przedstawił już w 1966 r. na gruncie socjologii polskiej Z. Gostkowski w pracy *Analiza „efektu panelowego” w badaniach wyborczych w Łodzi w r. 1961*. Badacz ten udowodnił, że udział respondenta w badaniu wpłynął na poziom zainteresowania wyborami, a także zmiany postaw, co udowodniły następane pomiary. Więcej informacji: Z. Gostkowski, *Analiza „efektu panelowego” w badaniach wyborczych w Łodzi w 1961 r.*, [w:] tegoż (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław – Warszawa – Kraków 1966, s. 282–310. Drugi efekt polega na tym, że respondenci celowo deklarują pewne istotne zmiany (choć w rzeczywistości one nie zaszły lub zająć w ogóle nie mogły – dotyczy to szczególnie danych dokumentalnych) albo – odwrotnie – dążą do zachowania spójności i stabilności swoich odpowiedzi (choć tu zmiana zaszła – np. w opiniach, postawach). Więcej informacji: A. Dyjas-Pokorska, *Badania trackingowe i panelowe*, [w:] Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork...*, s. 245–246. Na zaburzenie następnych pomiarów może wpłynąć dodatkowo sytuacja, gdy badani znają rezultaty pierwszego pomiaru: „Mogą one stanowić dla nich bardzo ważną informację, uświadamiającą im stopień rozbieżności między rzeczywistym stanem rzeczy a ich standardami i skłaniającą ich do działań zmierzających do ich uzgodnienia”; A. Sulek, *Eksperyment...*, s. 76–77. Więcej informacji na temat zjawiska pretestu i posttestu także w: F. Sztabiński, *Ocena jakości...*, s. 72; J. Brzeziński, *Metodologia badań psychologicznych*, Warszawa 2005, s. 314–316. Celem porównania warto także dodać, że w branży reklamowej zjawisko pretestu i posttestu podczas badań jest pożądane i jest źródłem cennych opinii dotyczących kampanii reklamowej. Więcej informacji: D. Maison, *O badaniach reklamy, czyli jak na podstawie badań przewidzieć skuteczność reklamy i ocenić skuteczność przeprowadzonej kampanii*, [w:] D. Maison, A. Noga-Bogomilski, *Badania marketingowe...*, s. 153–177.

- niezależność pomiaru, to znaczy pierwszy pomiar nie może wpływać na przebieg i wynik drugiego⁵⁶.

ANALIZA CZASU PRZEPROWADZANIA WYWIADÓW

Analiza czasu przeprowadzania wywiadów jest, zgodnie ze stanowiskiem F. Sztabińskiego, jednym z elementów oceny pracy ankierów i dotyczy poprawności realizacji wywiadu. Niemniej, jak wskazano wcześniej, sama kontrola ich pracy jest jedną z metod oceny wiarygodności wyników badania i jest jednym z czynników, które pomagają podjąć decyzję, czy informacje zebrane przez danego ankiera mogą się znaleźć w zbiorze, czy też powinny być usunięte. Należy wspomnieć, że stwierdzenie ewentualnych anomalii jest jedynie rekomendacją do głębszej kontroli, która pozwoli podjąć ostateczną decyzję. W obowiązującej praktyce badawczej, zgodnie z materiałami wewnętrznymi firm CBM Indicator i TNS OBOP, dodatkowymi elementami mogą być między innymi: braki numerów telefonów (zbierane do celów kontrolnych); odsetek poprawnych numerów telefonów (za niepoprawne przyjmuje się również te numery, które są cały czas wyłączone lub brak jest kompetentnego informatora); powtarzalność numerów telefonów⁵⁷ (pomiędzy różnymi badaniami u tego samego ankiera lub w tym samym regionie koordynatorskim⁵⁸).

Przy analizie czasu trwania badania bardzo pomagają programy komputerowe tworzące tzw. historię pracy ankiera⁵⁹. Wśród porównywanych parametrów, które pomagają w ocenie wiarygodności wyników, mogą się znaleźć: wielkość i typ miejscowości, w której był przeprowadzany wywiad, data jego przeprowadzenia, dzień tygodnia, godziny realizacji wywiadu i czas trwania oraz odstępy między wywiadami (dla badań CAPI czy PAPI)⁶⁰. Otrzymane rezultaty z dwu ostatnich parametrów porów-

⁵⁶ F. Sztabiński, *Ocena jakości...*, s. 72–73.

⁵⁷ Sprawdzane wyłącznie wtedy, gdy firma ma zgodę na przechowywanie numerów telefonów w dłuższej perspektywie czasowej. Zwykle numer telefonu jest bowiem usuwany bezpośrednio po zakończeniu projektu.

⁵⁸ Badania społeczne, ewaluacyjne i marketingowe realizowane przez wyspecjalizowane firmy – agencje badawcze – opierają się na współpracy z tzw. koordynatorami regionalnymi. Jest to tzw. model dwustopniowy – badanie zlecane jest koordynatorom, a dopiero oni przydzielają je ankierom (mało popularny jest model jednostopniowy, w którym pracownik agencji badawczej kontaktuje się bezpośrednio z ankierem z pominięciem koordynatora). Dzięki temu można w sposób istotny ograniczyć zatrudnienie w działach realizacji badań w agencjach.

⁵⁹ F. Sztabiński, *Kontrola realizacji...*, s. 356, 373.

⁶⁰ Opracowanie własne na podstawie: tamże, s. 373. W przypadku badań CATI porównywanych parametrów jest znacznie więcej, są to także parametry innego typu. Tutaj

nuje się do ogólnej średniej dla badania, a także do historii pracy innych ankierów i na tej podstawie formułuje się wnioski dotyczące jakości danych⁶¹. Do ilustracji tej procedury posłuży tabela 5. Dla uproszczenia przyjęto, że przykład dotyczy badania CAPI realizowanego w piątek. Data nie ma w tym wypadku znaczenia, ale założono, że tabela odnosi się do zwykłego okresu w roku, to jest przed piątkiem były cztery dni pracujące, a po tym dniu następują dwa dni wolne.

Tabela 5. Przykładowa historia pracy ankiera^a

Numer wywiadu	Typ miejscowości	Wielkość miejscowości (przedział liczby mieszkańców w tys.)	Godzina rozpoczęcia wywiadu	Godzina zakończenia wywiadu	Czas trwania rozmowy (w min.)	Odstęp między poszczególnymi wywiadami (w min.)
1	Wieś	–	9:12	9:37	25	–
2	Wieś	–	10:43	11:11	28	66
3	Wieś	–	13:07	13:45	38	116
4	Miasto	100–499	15:17	15:28	11	92
5	Miasto	100–499	15:41	16:09	28	13
6	Miasto	100–499	16:23	16:50	27	14
7	Miasto	100–499	17:14	17:39	25	24
8	Miasto	powyżej 500	19:21	19:43	22	102
9	Miasto	powyżej 500	20:14	20:47	33	31
10	Miasto	powyżej 500	21:19	21:48	29	32

^a Średnia długości przeprowadzanego wywiadu wyniosła 26 min. i 36 sek. (średnia dla całego badania: 31 min. 13 sek.), a średni odstęp między wywiadami – 54 min. 27 sek. (średnia dla całego badania podczas tego samego dnia: 61 min. 44 sek.).

Źródło: opracowanie własne na podstawie: Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork...*, s. 374.

można wymienić np. liczbę wykonanych połączeń podczas sesji w ogóle, liczbę wywiadów przerwanych, liczbę wywiadów przełożonych, średni czas trwania wywiadu, czas trwania wywiadu najkrótszego i najdłuższego. (Opracowanie własne na podstawie materiałów wewnętrznych firmy CBM Indicator). Badania CATI mają także tę przewagę, że w tym wypadku znacznie łatwiejsza jest możliwa bezpośrednia kontrola działań ankierów w czasie rzeczywistym. Kontrolerzy mogą bowiem podsłuchiwać wywiady w trakcie. Więcej informacji: Z. Sawiński, *Badania CATI i wywiady przez telefon*, [w:] Z. Sawiński, F. Sztabiński, P.B. Sztabiński (red.), *Fieldwork...*, s. 344–346.

⁶¹ Tamże, s. 373–375.

Historia pracy przedstawiona w tabeli 5 nakazuje bliżej przyjrzeć się pracy tego ankietera. O ile możliwe jest, że w trakcie swojego dnia pracy ankieter wykonał podróż ze wsi do dużego miasta, a średnie nie wzbudzają podejrzeń, o tyle zastanawiające są inne parametry jego pracy. Trudno uwierzyć, aby w średniej wielkości mieście (wywiady numer od 4 do 7) w porach dnia, gdy wiele osób jest niedostępnych, ponieważ pracują, ankieter mógł znaleźć respondentów w przedziale czasu tak krótkim w porównaniu do czasu na wsi. Ponadto budzą wątpliwości (choć nie tak duże) godziny zrealizowanych z sukcesem wywiadów w piątek w mieście liczącym powyżej 500 tys. mieszkańców.

POWTARZALNOŚĆ I CYKLICZNOŚĆ ODPOWIEDZI

Analiza powtarzalności i cykliczności odpowiedzi zalicza się także do nieterenowych metod kontroli⁶². F. Sztabiński zalicza ją też do kontroli wewnętrznej opierającej się na określeniu „częstotliwości pojawiania się określonych sekwencji odpowiedzi w kwestionariuszach ankieterów”⁶³, ale również na „porównywaniu prawdopodobieństw występowania określonych układów odpowiedzi”⁶⁴. Warto podkreślić, że tutaj metody statystyczne także mają zastosowanie (na przykład *data mining*⁶⁵). Z kolei patrząc od strony oszustw popełnianych przez ankieterów, dziwne wzorce odpowiedzi⁶⁶ mogą dotyczyć zarówno fałszerstw związanych z doborem respondenta⁶⁷, jak i nierzetelnego przeprowadzenia wywiadu (czego jednym z przejawów jest między innymi celowe ominięcie części pytań zawartych w kwestionariuszu⁶⁸).

W tabeli 6 przedstawiono różnicę między powtarzalnością a cyklicznością odpowiedzi. Oczywiście należy mieć na uwadze, że ten schemat jest bardzo uproszczony i dotyczy raczej badań PAPI.

⁶² Warto dodać, że F. Sztabiński w swojej publikacji *Ocena jakości...*, powołując się na Paula Lavrakasa, nazywa techniki nieterenowe *data analytic methods*. Oprócz wcześniej wspomnianej merytorycznej analizy wypełnionych kwestionariuszy do tej grupy zaliczają się też analizy opisu sytuacji wywiadu. Patrz: tenże, *Ocena jakości...*, s. 101.

⁶³ Tenże, *Kontrola realizacji...*, s. 356.

⁶⁴ Tenże, *Ocena jakości...*, s. 101.

⁶⁵ Z. Sawiński, F. Sztabiński, *Czy ankieterzy...*, s. 371.

⁶⁶ Tamże.

⁶⁷ Tamże, s. 367.

⁶⁸ Tamże, s. 368–373.

Tabela 6. Przykładowe rozkłady odpowiedzi na pytania dotyczące posiadania prawa jazdy i zachowań na drodze – różnica między powtarzalnością a cyklicznością (pogrubieniem zaznaczono odpowiedzi respondentów wprowadzone przez ankieterów, instrukcje dla ankietera – kursywą)

Pytania	Powtarzalność odpowiedzi			Cykliczność odpowiedzi		
1. Wpisz rok urodzenia respondenta (RRRR):	2003			1978		
2. Czy posiada Pan/Pani prawo jazdy? (ANKIETER: jeśli „tak” – pytanie 3. pozostaw puste i przejdź do pytania 4.)	tak	nie	odmowa	tak	nie	odmowa
3. A czy zamierza Pan/Pani zrobić je w ciągu najbliższego roku? (ANKIETER: w przypadku wątpliwości respondenta dodaj, że chodzi o sytuację, gdy respondent planuje zacząć kurs, ale też gdy jest jego uczestnikiem lub go zakończył i oczekuje na egzamin) (ANKIETER: jeśli „nie” – przejdź do następnej części badania, a kolejne pytania pozostaw puste)	tak	nie	odmowa	tak	nie	odmowa
4. Czy kiedykolwiek prowadził(-a) Pan/Pani samochód pod wpływem takiej ilości alkoholu, że Pan/Pani i/lub Pana/Pani otoczenie uznawało, że jest Pan/Pani pijany(-a)?	tak	nie	odmowa	tak	nie	odmowa
5. Czy kiedykolwiek prowadził(-a) Pan/Pani auto z prędkością wyższą o co najmniej 10 km/h niż to było dozwolone?	tak	nie	odmowa	tak	nie	odmowa
6. Czy kiedykolwiek otrzymał(-a) Pan/Pani mandat drogowy?	tak	nie	odmowa	tak	nie	odmowa
7. Czy kiedykolwiek otrzymał(-a) Pan/Pani przynajmniej jeden punkt karny?	tak	nie	odmowa	tak	nie	odmowa

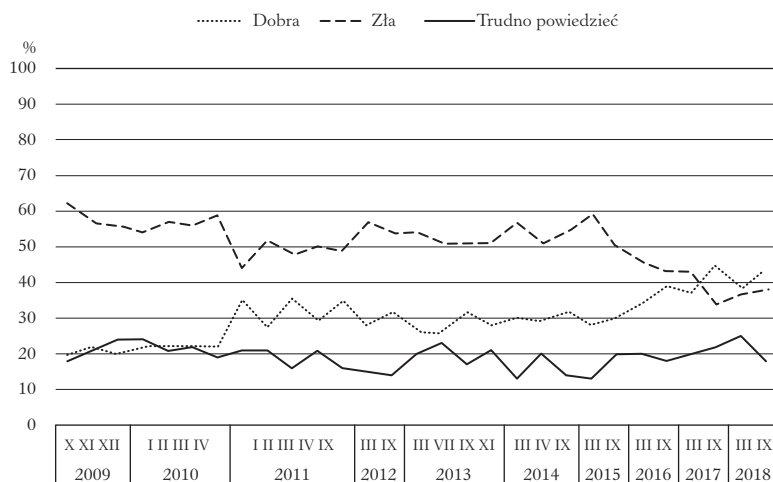
Źródło: opracowanie własne.

Gdyby odpowiedzi wprowadzone przez ankieterów (z wyjątkiem roku urodzenia) zamienić na liczby, w pierwszym przypadku (powtarzalności) byłby to układ: 1–1–1–1–1–1, a w drugim przypadku (cykliczności): 2–3–1–2–3–1. Powtarzalność cechuje zatem występowanie tych samych odpowiedzi, a cykliczność regularne występowanie tego samego układu (w omawianym przykładzie: 2–3–1). O fałszerstwie, oprócz zaobserwowanych anomalii, świadczy także brak wewnętrznej logicznej spójności odpowiedzi.

SEZONOWOŚĆ W BADANIACH POWTARZALNYCH I CIĄGŁYCH

Badania powtarzalne i ciągłe są szczególnym typem wywiadu ilościowego, który realizowany jest na dużej próbie, a sam pomiar realizowany jest wielokrotnie przy użyciu takiego samego lub nieco modyfikowanego narzędzia. Przy każdym pomiarze próba jest inna, ale dobierana w ten sam sposób i reprezentuje tę samą populację generalną⁶⁹. Jak pisze Anna Dyjas-Pokorska: „trackingi [badania powtarzalne i ciągłe – aut.] służą do badania dynamiki zachowań, postaw, preferencji i wyborów konsumenc- kich lub społecznych i do śledzenia trendów, a więc przewidywania, jakie zmiany w mierzonych zjawiskach mogą nastąpić w najbliższej i dalszej perspektywie”⁷⁰.

Ze względu na powtarzalność pomiaru możliwe jest zaobserwowanie, czy oraz kiedy (w jakich warunkach) występują regularne, powtarzalne zmiany dotyczące badanego zjawiska. Za przykład niech posłuży wykres opracowany na podstawie badań statutowych Centrum Badania Opinii Społecznej (rys. 4).

Rysunek 4. Oceny działalności Zakładu Ubezpieczeń Społecznych w latach 2009–2018^a

^a Połączenia między kolejnymi punktami pomiarowymi służą jedynie zilustrowaniu wzrostów lub spadków ocen.

Źródło: opracowanie własne na podstawie raportów Centrum Badania Opinii Społecznej: *Opinie o działalności prezydenta, parlamentu, ZUS, ABW i CBA*, komunikat z badań BSS/166/2009; *Oceny instytucji publicznych*, komunikat z badań BS/128/2012; *Oceny działalności instytucji publicznych*, komunikat z badań nr 121/2018.

⁶⁹ A. Dyjas-Pokorska, *Badania trackingowe...*, s. 237–238.

⁷⁰ Tamże.

Jak widać na wykresie, w latach 2015–2018 oceny działalności ZUS we wrześniu ulegają względem marca łagodnej poprawie. Jednym z uzasadnień występowania sezonowości może być fakt, że pomiędzy marcem a kwietniem do emerytów i rencistów docierają informacje o rewaloryzacji rent i emerytur⁷¹, a we wrześniu respondenci mają już za sobą kilka wypłat świadczenia w nowej wysokości⁷². Ponadto pod koniec każdego roku i na początku roku następnego przedsiębiorcy (oraz osoby pracujące) dowiadują się o wysokości składek wpłacanych do Funduszu Ubezpieczeń Społecznych⁷³. Może to dodatkowo tłumaczyć łagodne spadki między wrześniem a marcem następnego roku (wyjątek stanowi przełom roku 2015 i 2016, kiedy nastąpiło zaburzenie tego trendu). Należy także zwrócić uwagę na rok 2012, w którym pomiary zostały dokonane w takim samym interwale czasowym, wspomniane zaś wcześniej zmiany również są widoczne. Sezonowość widać również w 2011 roku. Jednakże, ze względu na nierównomierność rozmieszczenia punktów pomiarowych w czasie, wnioskowanie nie jest dokładne. Warto przy tym pamiętać, że im więcej punktów pomiarowych, tym dokładniejszy będzie wyznaczony trend.

Weryfikacja danych – kilka praktycznych przykładów

Analiza zjawisk otaczającego świata może zostać dokonana z użyciem wielu różnych metod statystycznych. Niektóre z nich dostarczają wartościowej poznawczo informacji i nie wymagają korzystania z zaawansowanych instrumentów. Istnieją narzędzia, które w prosty sposób umożliwiają sprawdzenie, czy zbiór danych, który jest w naszym posiadaniu, bądź jakieś jego elementy nie zostały sfalszowane. Do pierwszej grupy należy rozkład normalny Carla Friedricha Gaussa oraz rozkład zdarzeń

⁷¹ Zgodnie z art. 88 ust. 1 ustawy z 17 grudnia 1998 r. o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych (tekst jedn. Dz.U. 2018, poz. 1270, z późn. zm.) – stan prawny na dzień 8.01.2019.

⁷² Z kolei jednym z argumentów przeczących temu wyjaśnieniu jest fakt, że w połowie każdego roku Polacy otrzymują od ZUS prognozę swojej przyszłej emerytury. Są one zauważalnie niższe względem nawet obecnych świadczeń emerytalnych.

⁷³ *Najniższa podstawa wymiaru składek oraz kwoty składek na ubezpieczenia społeczne w 2019 r.*, http://www.zus.pl/o-zus/komunikaty/-/publisher/komunikat/1/najnizsza-podstawa-wymiaru-skladek-oraz-kwoty-skladek-na-ubezpieczenia-spoleczne-w-2019-r_/2136916 (dostęp: 10.01.2019); *Wysokość składek na ubezpieczenia*, <http://www.zus.pl/baza-wiedzy/skladki-wskazniki-odsetki/skladki/wysokosc-skladek-na-ubezpieczenia-spoleczne> (dostęp: 10.01.2019).

rzadkich Siméona Denisa Poissona. Weryfikacja elementów wchodzących w skład zbioru danych może odbyć się z użyciem testu Ulricha Grafa, kryterium Williama Chauveneta bądź z zastosowaniem prawa Franka Benforda.

WERYFIKACJA ZBIORÓW DANYCH – ROZKŁAD GAUSSA I ROZKŁAD ZDARZEŃ RZADKICH POISSONA

Rozkład normalny, choć powszechnie używany w nauce, pozostaje niedoceniony w pracy analityków informacji. Nazwę „rozkład normalny” wprowadził w 1889 roku brytyjski antropolog, genetyk i statystyk Francis Galton. Niepośledni wkład w badania nad tym fenomenem wniósł niemiecki matematyk, fizyk i astronom Carl Friedrich Gauss – to od jego nazwiska właśnie wzięła się nazwa krzywej dzwonowatej, opisującej między innymi liczne zjawiska występujące w przyrodzie czy cechy charakteryzujące organizmy, a także, co bardzo istotne, rozkład błędów pomiaru.

Prace C.F. Gaussa w tej materii zostały zapomniane przez naukowców na długie lata. Popularność i wszechstronność zastosowań rozkładu normalnego przywrócił francuski matematyk, astronom, geodeta i fizyk Pierre Simon de Laplace⁷⁴. To właśnie on zetknął się z pracami C.F. Gaussa w 1810 roku, kiedy przedstawił Akademii Nauk pracę, w której udowodnił tak zwane centralne twierdzenie graniczne, zgodnie z którym prawdopodobieństwo, iż suma dużej liczby niezależnych czynników losowych przyjmie wartość z danego zakresu, jest takie jak przewiduje rozkład normalny. Czytając prace Gaussa, Laplace zdał sobie sprawę, że mógłby z nich skorzystać i poprawić własne wyniki oraz że jego uzupełniona w ten sposób praca dostarcza lepszych argumentów niż te użyte przez samego Gaussa na poparcie tezy, że rozkład normalny stanowi tzw. uniwersalne prawo błędu. P.S. Laplace szybko opublikował krótki dodatek do swojej wcześniejszej pracy. Centralne twierdzenie graniczne oraz prawo wielkich liczb są do dnia dzisiejszego dwoma najsłynniejszymi twierdzeniami w rachunku prawdopodobieństwa.

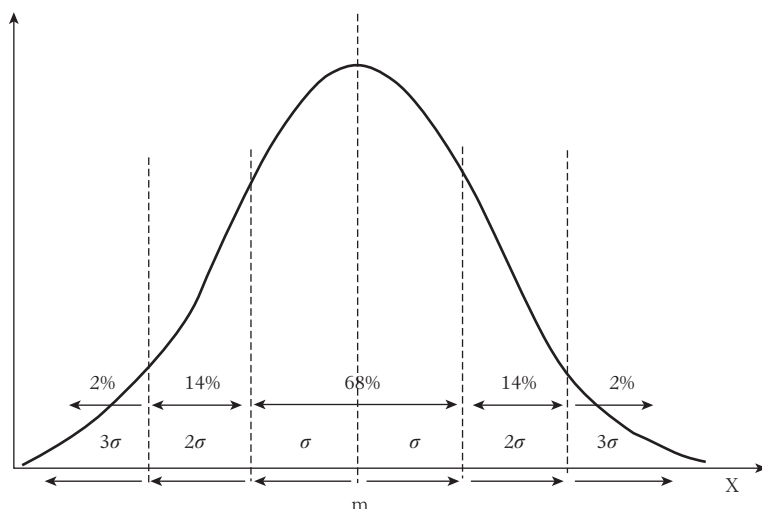
Wśród prekursorów współczesnej statystyki bardzo często wspomniany jest żyjący w XIX wieku belgijski astronom, matematyk, meteorolog, socjolog i kryminolog Lambert Adolphe Jacques Quételet, który przeprowadził obszerne badania, aby wykazać statystyczną regularność pewnych cech czy faktów (jak liczba urodzeń, zgonów czy popełnianych

⁷⁴ L. Młodinow, *Matematyka niepewności. Jak przypadki wpływają na nasz los*, przekł. P. Strzelecki, Warszawa 2009, s. 168–170.

przestępstw). W poszukiwaniu danych potwierdzających występowanie rozkładu normalnego pomogło mu pewne szkockie czasopismo, które opublikowało dane dotyczące obwodu klatki piersiowej oraz wzrostu ponad 5000 żołnierzy odbywających służbę w różnych szkockich pułkach. Dane te pozwoliły mu wykazać, że zmienność cech żołnierzy jest tego samego rodzaju co opisywana rozkładem normalnym. Dalsze prace uczonego w tej materii zaowocowały opracowaniem wykorzystywanego do dnia dzisiejszego wskaźnika masy ciała, zwanego wskaźnikiem Quételeta lub BMI (*Body Mass Index*).

Krzywa Gaussa pokazuje, z jakim prawdopodobieństwem w dowolnej populacji, w tym ludzkiej, występują poszczególne wartości danej cechy. Mogą to być na przykład: iloraz inteligencji, wzrost, zarobki itd. Wierzchołek krzywej opisuje średnią wartość występowania cechy, a wielkość oznaczona symbolem greckim σ – odchylenie standardowe. Ta ostatnia informuje o tym, jak szeroko wartości badanej cechy rozrzucone są wokół średniej. Kształt krzywej zmienia się w zależności od wartości odchylenia standardowego – im jest ono mniejsze, tym krzywa jest bardziej smukła. Dla dużych wartości odchylenia standardowego kształt krzywej zmienia się na bardziej rozłożysty. W przypadku rozkładu normalnego 68,2% wartości cechy leży w odległości mniejszej lub równej σ od wartości średniej (patrz rys. 5 – pas $(\bar{x} - \sigma; \bar{x} + \sigma)$), 95,4% wartości cechy leży w odległości 2σ od średniej, a 99,6% w odległości 3σ .

Rysunek 5. Parametry rozkładu normalnego



Źródło: opracowanie własne.

Praktyczną analityczną przydatność rozkładu normalnego prześledzono w następującej egzemplifikacji. Krzywa normalna między innymi opisuje rozkład ilorazu inteligencji w populacji mieszkańców Europy. Standardowo przyjmuje się, że średnia wartość tej cechy wynosi 100 IQ, a odchylenie standardowe – 15. Zatem 68% populacji to osoby mające średni iloraz inteligencji (pas pomiędzy 85 a 115 IQ). Ludzie o inteligencji wyższej niż przeciętna – mający IQ powyżej 116, ale poniżej 130 – stanowią blisko 14% populacji. Tyle samo jest osób mało inteligentnych mających IQ między 70 a 84. Skrajne wartości cechy znajdujące się po obu stronach wartości średniej, a oddalone od niej o odległość 3σ , to ludzie uznawani za upośledzonych umysłowo – IQ poniżej 70 (lewa strona krzywej) lub ponadprzeciętnie inteligentnych – IQ powyżej 130 (prawa strona krzywej). Podsumowując: na 100 Europejczyków dwóch będzie upośledzonych umysłowo, 14 – mało inteligentnych, 68 – średnio inteligentnych, 14 – o ponad przeciętną inteligencję i 2 genjuszy.

Kolejnym jakże istotnym, z punktu widzenia statystyki, jest rozkład Poissona, opisujący tzw. rozkład zdarzeń rzadkich, znany także jako prawo małych liczb Poissona. Jest wykorzystywany wszędzie tam, gdzie prawdopodobieństwo wystąpienia danego zjawiska jest bardzo małe, niemalże równe zero, natomiast liczba jednostek, którym można je przypisać, jest bardzo duża, zbiegająca nawet do nieskończoności. Twórcą tego rozkładu jest Siméon Denis Poisson, francuski matematyk i fizyk, który w pracach nad badaniem prawdopodobieństwa wydania danego wyroku sądowego w sprawach cywilnych i karnych wprowadził teorię pozwalającą przewidzieć występowanie określonego zdarzenia w wyznaczonym przedziale czasu⁷⁵. Badaczem, który wniósł duży wkład w rozwój tej teorii, był rosyjski matematyk polskiego pochodzenia Władysław Bortkiewicz. Stąd niekiedy w literaturze można się spotkać z inną nazwą teorii – prawem małych liczb Poissona–Bortkiewicza. W. Bortkiewicz znalazł zastosowanie rozkładu Poissona w szacowaniu wśród kawalerzystów pruskich korpusów liczby zgonów spowodowanych kopnięciem konia⁷⁶. Rozkład ten jest bardzo często wykorzystywany w medycynie, finansach oraz ubezpieczeniach do przewidywania liczby zachorowań na bardzo rzadką chorobę, wykrycia systemowych błędów pojawiających się podczas księgowania transakcji finansowych czy do oszacowania prawdopodobieństwa, że na

⁷⁵ D. Mider, A. Marcinkowska, *Analiza danych ilościowych dla politologów. Praktyczne wprowadzenie z wykorzystaniem programu GNU PSPP*, Warszawa 2013, s. 268–269.

⁷⁶ Więcej na temat W. Bortkiewicza oraz jego dokonań w: J. Schumpeter, *Ladislav von Bortkiewicz*, „*Economic Journal*” 1932, nr 42, s. 338–340; P.A. Samuelson, *Resolving a Historical Confusion in Population Analysis*, „*Human Biology*” 1976, nr 48, s. 559–580.

początku trwania ubezpieczenia osoba, która dokonała zakupu polisy, dozna trwałego uszczerbku na zdrowiu.

Rozkład Poissona jest przykładem rozkładu dyskretnego (skokowego) i jest szczególnym przypadkiem rozkładu Jakoba Bernoulliego, definiowanego za pomocą dwóch parametrów: λ oraz k , gdzie $\lambda = n * p$ (n – liczba obserwacji, p – prawdopodobieństwo wystąpienia zdarzenia), a k jest liczbą sukcesów wystąpienia zdarzenia o prawdopodobieństwie p . Rozkład Poissona określony jest następującym wzorem:

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}$$

gdzie e jest liczbą Nepera, równą w przybliżeniu 2,72.

Ciekawą własnością tego rozkładu jest to, że wartość oczekiwana jest równa wariancji i wynosi λ . W przypadku, kiedy λ przybiera bardzo duże wartości, rozkład Poissona jest z dobrą dokładnością przybliżany przez rozkład Gaussa z tą samą wartością średnią i odchyleniem standardowym. Nie jest więc błędem stosowanie rozkładu Gaussa zamiast rozkładu Poissona w przypadkach, kiedy iloczyn liczby obserwacji i prawdopodobieństwa wystąpienia danego zdarzenia przyjmuje dużą wartość.

W statystyce istnieje wiele testów sprawdzających normalność rozkładu. Badanie, czy rozkład zmiennej jest zbliżony do rozkładu normalnego, jest podstawą do stosowania dalszych metod statystycznych, takich jak: analiza wariancji, korelacja r-Pearsona, regresja wieloraka itd. W dalszej części skupimy się na jednym z najczęściej stosowanych testów opracowanym w 1965 roku przez dwóch statystyków: Samuela Shapiro i Martina Wilka, który – jak wykazała analiza porównawcza metodą Monte Carlo – ma największą moc spośród innych testów sprawdzających normalność rozkładu. Żeby zapoznać czytelnika z łatwością stosowania oraz interpretacją tego testu (bez znajomości zaawansowanej wiedzy statystycznej czy matematycznej), posłużmy się następującym przykładem. Załóżmy, że w pewnym bardzo małym mieście urząd skarbowy pozyskał informacje dotyczące wysokości zarobków swoich mieszkańców z deklaracji podatkowych, które złożyli. Naczelnik urzędu jest osobą bardzo podejrzliwą i chce sprawdzić testem Shapiro–Wilka, czy uzyskane dane nie zostały przez podatników zafałszowane. Wprowadził wyniki do programu SPSS⁷⁷ i wykonał kolejno następujące polecenia:

⁷⁷ SPSS (Statistical Package for the Social Sciences) – płatne oprogramowanie do statystycznej analizy danych. Najczęściej wykorzystywane jest w badaniach naukowych, rynku i opinii, badaniach epidemiologicznych. Typowa praca z SPSS jest pracą z oknami

w menu wybrał Analiza → Opis statystyczny → Eksploracja, w oknie dialogowym w miejscu Zmienne zależne wprowadził zmienną „zarobki”, a w kolejnym kroku w menu Wykresy zazaczył opcję Wykresy normalności z testami oraz Histogram. W efekcie otrzymał różne tabelki i wykresy. Najważniejszą z nich jest ta o nazwie Testy normalności rozkładu, z niej można odczytać, czy podejrzenie o zafałszowanie danych jest słuszne. W tabeli 7 prezentujemy wynik obliczeń, jaki otrzymał nasz naczelnik.

Tabela 7. Testy normalności rozkładu (obrazujące przykładowe działania weryfikacji poprawności danych)

	Kołmogorow-Smirnow ^a			Shapiro-Wilk		
	Statystyka	df	Istotność	Statystyka	df	Istotność
zarobki	,081	77	,200*	,964	77	,027

* dolna granica rzeczywistej istotności

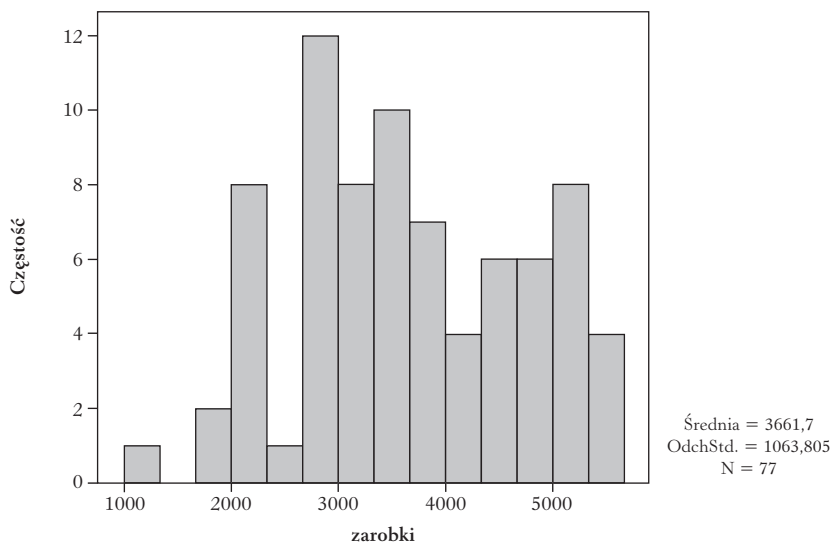
^a z poprawką istotności Lillieforsa

Źródło: opracowanie własne.

Tabela zawiera wyniki dwóch testów, ale naczelnika interesuje tylko część prezentująca wynik testu Shapiro–Wilka. Kluczową informacją jest wartość współczynnika istotności. Jeśli jest ona niższa niż 0,05, to możemy sądzić, że dane, które poddaliśmy testom, nie spełniają warunku normalności rozkładu. W naszym przykładzie wartość ta wynosi 0,027 i jest mniejsza niż 0,05, a zatem podejrzenie o zafałszowaniu zarobków mieszkańców miasta zostało potwierdzone, a ujawnia ten fakt otrzymany histogram. Już „na oko” widać, że gdybyśmy połączyli środki każdego z wierzchołków słupków, to linia je łącząca nie będzie w żaden sposób przypomiwała krzywej Gaussa.

dialogowymi i kreatorami graficznymi. Od 1968 r. w sposób ciągły jest rozbudowywane o nowe moduły i metody analityczne. Darmową alternatywą dla SPSS jest GNU PSPP, w którym stosowany język dąży do zgodności ze swoim płatnym odpowiednikiem. Obecnie na rynku jest wiele programów wspomagających przeprowadzanie analiz statystycznych. Autorzy proponują wykorzystanie GNU PSPP nie tylko ze względu na to, że jest bezpłatny, ale również ze względu na wygodę i intuicyjne jego wykorzystanie.

Rysunek 6. Przykładowy histogram prezentujący rozkład zmiennej



Źródło: opracowanie własne.

Równie szybkie i proste jest sprawdzenie, czy uzyskane w toku pomiarów dane przyjmują rozkład Poissona. Wykorzystamy w tym celu test Kołmogorowa–Smirnowa, którego nazwa pochodzi od nazwisk dwóch słynnych rosyjskich matematyków: Andrieja Kołmogorowa oraz Nikołaja Smirnowa. Podobnie jak przy rozkładzie normalnym posłużmy się następującym przykładem. Pewien pracownik dużej firmy otrzymuje dodatkową premię – za każdego klienta, z którym podpisze długotrwałą umowę na wykorzystanie pewnego sprzętu. Z informacji uzyskanych od analityków tej firmy wynika, że podczas jednego dnia pracy podpisywane są średnio cztery umowy z nowymi klientami. Pracownik złożył kwartalny raport z wyników swojej pracy. Jego przełożony przypuszcza, że pracownik zawyżył liczbę umów, które podpisał w tym czasie, celem wymuszenia na pracodawcy wyższej premii. Wykaz dokładnej liczby codziennych sprzedaży został poddany testowi Kołmogorowa–Smirnowa. Wykonano następujące polecenia w programie SPSS: Analiza → Testy nieparametryczne → Testy tradycyjne → K-S dla jednej próby, następnie wybrano rodzaj testu → Poissona. Otrzymano dwie tabele, z których najważniejsza to test Kołmogorowa–Smirnowa dla jednej próby. Podobnie jak w poprzednim przykładzie, informacją potwierdzającą lub zaprzeczającą przypuszczeniom zafałszowania liczby umów jest wartość współczynnika istotności asymptotycznej. Jeśli jest on większy od 0,05, to nasza zmienna przyjmuje rozkład Poissona, w przeciwnym wypadku możemy

stwierdzić, że dane takiego rozkładu nie przypominają. W naszym przykładzie wartość ta wynosi 0,037, jest zatem mniejsza niż 0,05. Przełożony może zatem śmiało stwierdzić, że jego podwładny dokonał zafałszowania wyników swojej pracy.

Rysunek 7. Test Kołmogorowa–Smirnowa dla jednej próby

		VAR00002
N		92
Parametr rozkładu	Średnia	5,6413
Największe różnice	Wartość bezwzględna	,147
	Dodatnia	,147
	Ujemna	-,047
Z Kołmogorowa–Smirnowa		1,413
Istotność asymptotyczna (dwustronna)		,037

Źródło: opracowanie własne.

WERYFIKACJA POSZCZEGÓLNYCH ELEMENTÓW ZBIORU DANYCH – KRYTERIUM CHAUVENETA, TEST GRAFA, PRAWO BENFORDA

Wymienione tu testy umożliwiają identyfikację odstających obserwacji w zbiorze danych. Czy odbiegająca wartość jest spowodowana błędem pomiaru i należy ją odrzucić, czy też uzyskany wynik może być odzwierciedleniem jakiegoś ważnego efektu, a jego pojawienie się nie jest żadną pomyłką? Zwykle niemożliwe jest ustalenie zewnętrznych przyczyn powstania anomalnego wyniku, a decyzja o pozostawieniu go w zbiorze danych lub usunięciu należy do badacza. Rozstrzygnięcie dylematu nie jest sprawą prostą, zwłaszcza w sytuacjach, kiedy w grę wchodzi bardzo precyzyjne wykonanie analiz, na podstawie których mają być podjęte ważne lub strategiczne decyzje, na przykład o wprowadzeniu nowego leku czy w sprawie produkcji urządzeń wysokiej precyzji.

Opracowano kilka testów bazujących na porównaniach różnic od wartości średniej i odchylenia standardowego, które pozwalają na wykrycie odstających obserwacji przy założeniu, że pomiar wielkości ma rozkład normalny. Do najważniejszych z nich i najczęściej stosowanych należą: kryterium Williama Chauveneta, test Ulricha Grafa, test Franka Grubbsa oraz kryterium Benjamina Peirce'a. W literaturze przedmiotu wszystkie należą do grupy tzw. testów na błąd gruby⁷⁸. W niniejszym artykule

⁷⁸ W literaturze możemy spotkać się z podziałem błędów pomiaru na trzy kategorie: błędy przypadkowe (niepowtarzające się, przyjmujące wartości całkowite, wynikające z nie-

szczegółowo zostaną omówione pierwsze dwa. Przed przystąpieniem do realizacji testu z wykorzystaniem kryterium Chauveneta należy mieć pewność, że każdy z następujących trzech warunków został spełniony: wartości zmiennej muszą mieć rozkład normalny; tylko jeden z wyników pomiaru może znacznie odbiegać od pozostałych; mamy niemalże pewność, że podejrzany wynik jest przejawem błędu, a nie odzwierciedleniem jakiegoś ważnego efektu.

Obliczeń dokonuje się według następującego algorytmu⁷⁹:

- obliczamy średnią \bar{x} oraz wariancję s z całości próby;
- liczymy odchylenie otrzymanego wyniku od wartości średniej ze

wzoru: $t = \frac{|x - \bar{x}|}{s}$ gdzie x , to nasza odstająca wartość;

- liczymy prawdopodobieństwo, że nasz podejrzany wynik będzie oddalony od wartości średniej o nie mniej niż $(t * s)$ za pomocą wzoru: $p = 2 - 2\varphi(t)$, gdzie $\varphi(t)$ jest wartością dystrybuanty rozkładu normalnego w punkcie t (odczytujemy ją z tablic rozkładu normalnego);
- obliczamy iloczyn $(n * p)$;
- sprawdzamy, czy $(n + p) < 0,5$. Jeśli tak, podejrzany wynik należy odrzucić, jeśli nie, to należy pozostawić go w zbiorze wyników.

Jak już zostało wspomniane, kryterium Chauveneta stosujemy w przypadku, gdy mamy do czynienia tylko z jednym odstającym wynikiem pomiaru. Co zatem należy zrobić, gdy jest ich więcej? Załóżmy na początek, że mamy dwa podejrzane wyniki: x_1 oraz x_2 , przy czym x_2 jest bardziej odległe od średniej niż x_1 . W pierwszej kolejności powinniśmy zastosować kryterium Chauveneta do x_1 . Jeśli spodziewana liczba pomiarów równie odległych jest mniejsza niż jeden, powinniśmy odrzucić obie wartości. Jeśli przeciwnie – liczba ta jest większa niż jeden, to stosujemy kryterium używając x_2 . W przypadku, kiedy uzyskamy liczbę pomiarów równie odległych od średniej mniejszą niż 0,5, to x_2 należy odrzucić. Po odrzuceniu dowolnego pomiaru, który nie spełnia kryterium, należy ponownie obliczyć \bar{x} i s , używając tylko pozostałych danych. Nowa wartość s będzie mniejsza niż pierwotna i dla niej więcej pomiarów mogłoby

kontrolowanych podczas pomiaru czynników – np. zmiany napięcia sieci elektrycznej czy też ograniczonej dokładności obserwacji), błędy grube (duże błędy spowodowane nieuwagą osoby dokonującej pomiaru) oraz błędy systematyczne (powtarzające się, powodowane przez czynniki działające w jednakowy sposób, w czasie wielokrotnego powtarzania tego samego pomiaru).

⁷⁹ Patrz także: J.R. Taylor, *Wstęp do analizy błęd pomiarowego*, przekł., A. Babiński, R. Bożek, Warszawa 2018, s. 200–202.

nie spełniać kryterium Chauveneta. Wielu naukowców podchodzi jednak bardzo sceptycznie do stosowania kryterium z przeliczonymi wartościami średniej oraz wariancji. Dlatego warto zastosować inne metody wykrywania odstających obserwacji, kiedy w wyniku pomiaru pojawiło się ich więcej niż dwie.

Prostszym i szybszym w obliczeniach algorytmem oceny, czy daną wartość należy pozostawić w zbiorze, czy odrzucić, jest test Grafa. Należy jednak pamiętać, że test ten jest mniej „czuły”⁸⁰ na wyszukiwanie danych odbiegających niż kryterium Chauveneta, co związane jest w wykorzystaniem w przypadku tego drugiego precyzyjniejszych narzędzi matematycznych. Zaleca się, aby po realizacji testu Grafa otrzymane wyniki dodatkowo poddać analizie kryterium Chauveneta.

Test Grafa przeprowadza się według następującego algorytmu⁸¹:

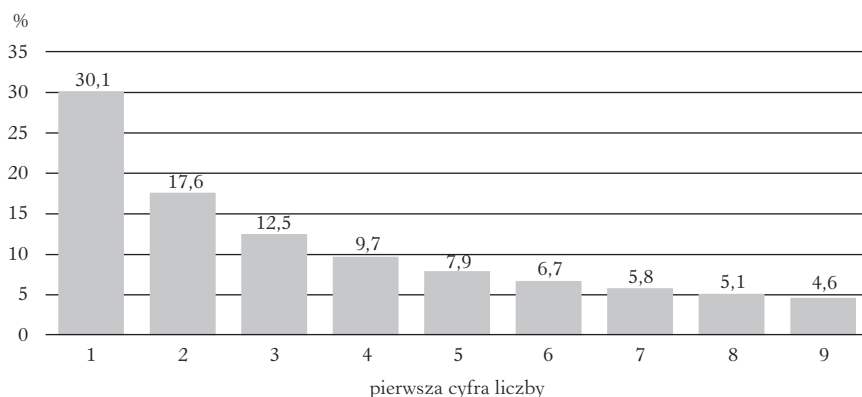
- liczymy wartość średnią \bar{x} oraz wariancję s z wyłączeniem podejrzanej obserwacji;
- wyliczone wartości podstawiamy do wzoru $(\bar{x} - 4s; \bar{x} + 4s)$ opisującego przedział dopuszczalnych wyników;
- sprawdzamy, czy nasza podejrzana wartość mieści się w tym przedziale – jeśli tak, to należy ją pozostawić w zbiorze, a w przeciwnym przypadku należy ją odrzucić.

Kolejny rodzaj testu to prawo Benforda. Jednym z pierwszych odkrywców ciekawej zależności, jaką wykazują liczby wiodące, to jest pierwsze cyfry występujące w wielu zbiorach danych liczbowych, był astronom i matematyk Simon Newcomb. Szukając w bibliotece United States Naval Observatory materiałów do swojej nowej pracy, zauważył, że strony ksiąg z tablicami logarytmicznymi są bardziej zużyte na początkowych niż na końcowych kartach. Przyczyną był fakt zapotrzebowania przez badaczy częściej na logarytmy liczb zaczynających się od 1 niż innych. W 1881 roku S. Newcomb sformułował i opublikował w „American Journal of Mathematics” hipotezę, że częstość występowania cyfr wiodących ma rozkład procentowy taki, jak na rysunku 8.

⁸⁰ M. Słowik, M. Bartkowiak, *Elementy statystycznej analizy wyników pomiarów na przykładzie badań wybranych cech mieszanek mineralno-asfaltowych*, „Drogownictwo” 2016, nr 7–8, s. 247–253.

⁸¹ Patrz także: J.R. Taylor, *Wstęp do analizy...*, s. 200–202.

Rysunek 8. Rozkład częstości występowania liczb znaczących



Źródło: opracowanie własne.

Liczba 1 występuje w 30,1% przypadków, 2 – w 17,6%, 3 – w 12,5%. Spadek częstości występowania jest tak gwałtowny, że jedynki są niemalże siedem razy bardziej prawdopodobne od dziewiątek.

Rozkłady procentowe S. Newcomba oparte są na logarytmach⁸². Wywnioskował on, że prawdopodobieństwo tego, iż dana liczba zaczyna się na cyfrę a , wynosi $[\log(a+1) - \log a]$. Jego odkrycie pozostało jednak bez echa najprawdopodobniej dlatego, że nie przedstawił ścisłego dowodu, co w świecie matematyków pozwala je traktować jedynie jako ciekawostkę.

Niemalże pół wieku później, w 1938 roku, tego samego odkrycia (bez znajomości wcześniejszej pracy S. Newcomba), również na podstawie tablic logarytmicznych, dokonał Frank Benford, fizyk zatrudniony w General Electric w Nowym Jorku. F. Benford nie ograniczył się jednak do tablic logarytmicznych. Przeanalizował występowanie cyfr wiodących pochodzących między innymi z tabel populacji amerykańskich miast, adresów pierwszych kilkuset osób, których dane zawarte były w *American Men of Science*, tablic ciężarów atomowych pierwiastków, zestawień powierzchni rzek i danych dotyczących wyników meczy baseballowych. Dla niemalże wszystkich zestawów liczbowych rozkład liczb wiodących był bliski oczekiwanego, z błędami na poziomie kilku dziesiątych procenta. F. Benford postawił hipotezę, że stanowi to przejaw uniwersalnego

⁸² A. Bellos, *Alex po drugiej stronie lustra. Jak liczby odzwierciedlają życie, a życie odzwierciedla liczby*, przekł. M. Krośniak, Warszawa 2018, s. 52–53.

prawa, które nazwał prawem anomalnych liczb⁸³. Nazwa ta się jednak nie przyjęła, obecnie prawidłowość ta zwana jest prawem Benforda.

Metoda sprawdzania zgodności pierwszych cyfr z prawem Benforda wykorzystywana jest obecnie do wykrywania manipulacji liczbami wszędzie tam, gdzie sensowne jest założenie, że prawo to powinno być spełnione.

Przykłady praktycznego zastosowania prawa Benforda są liczne. Do najślynniejszych należy sprawa „kreatywnej księgowości” z 1992 roku. Postawiono wówczas w akt oskarżenia Jamesa Nelsona, głównego księgowego Arizona State Treasurer, zarzucając mu defraudację niemalże 2 mln dolarów. W poczet dowodów przeciwko oskarżonemu została zaliczona analiza finansów z wykorzystaniem prawa Benforda, która stanowiła kluczowy argument pozwalający na wydanie skazującego wyroku⁸⁴. Walter Mebane, politolog z University of Michigan, użył prawa Benforda do udowodnienia manipulacji wynikami wyborów prezydenckich w Iranie w 2009 roku. Analizy wykazały, że głosy oddane na urzędującego prezydenta Mahmuda Ahmadineżada wyraźnie odbiegały od testu cyfr wiodących, a oddane na jego przeciwnika, Mira Hosseina Musawiego, rozbieżne nie były⁸⁵. Naukowcy wykorzystują również omawiane prawo jako narzędzie diagnostyczne. Malcolm Sambridge z Australian National University przeanalizował zapisy dwóch różnych sejsmografów, które zarejestrowały trzęsienie ziemi w Indonezji w 2004 roku – z Peru i z Australii. Wyniki podawane przez pierwszy z wymienionych były zgodne z prawem Benforda, drugiego zaś nie. M. Sambridge wyciągnął wniosek, że te drugie dane te musiały być zaburzone przez wstrząsy sejsmiczne o niewielkiej sile w rejonie Canberry. W tym przypadku test cyfr wiodących ujawnił trzęsienie ziemi, które w innym przypadku pozostałoby niezauważone⁸⁶.

Dowód prawa Benforda nie jest matematycznie prosty, ale zastosowanie ma wręcz banalne. Aby zrozumieć dowód, należy znać teorię

⁸³ F. Benford, *The Law of Anomalous Numbers*, „Proceedings of the American Philosophical Society” 1938, nr 78(4), s. 551–572.

⁸⁴ S.A. Barnes, *Identifying Fraud Can Be as Easy as 1, 2, 3: Applying Benford's Law to Forensic Analyses and Investigations*, <https://files.constantcontact.com/85e69272601/6ec4247f-14c0-49cf-92a6-674a5a4c37c2.pdf> (dostęp: 19.01.2019).

⁸⁵ W.R. Mebane, Jr., K. Kalinin, *Comparative Election Fraud Detection*, APSA 2009 Toronto Meeting Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450078, 2009 (dostęp: 19.01.2019).

⁸⁶ M. Sambridge, H. Tkalčić, P. Arroucau, *Benford's Law of First Digits: From Mathematical Curiosity to Change Detector*, „Asia Pacific Mathematics Newsletter” 2011, nr 1, s. 1–5.

ergodyczną, łączącą teorię prawdopodobieństwa z fizyką statystyczną. Ciekawą rzeczą jest natomiast właściwość niezmienniczości skalowania tego prawa. Jeśli jakieś dane finansowe wyrażone w złotówkach spełniają test cyfr wiodących, to będzie on również spełniony po przeliczeniu złotych na dolary czy euro. Aby samemu się przekonać o tym, że niemalże wszystkie zjawiska, jakie zachodzą w naszym otoczeniu, spełniają prawo Benforda, proponujemy wziąć zwykłą gazetę. Policzywszy wiodące cyfry, każdy jest w stanie się przekonać, że ich rozkład ułoży się wzdłuż opadającej skali z jedynek znacznie przeważającą pod względem częstości wobec pozostałych cyfr. Prawo Benforda zapoczątkowało odkrycie serii innych praw i zależności, takich jak prawo George'a Kingsley'a Zipfa oraz prawo Vilfreda Pareto.

Opinie ekspertów

W artykule przedstawiono wiele różnych metod weryfikacji danych. Metod zarówno niestatystycznych, zdroworozsądkowych, jak i tych wywodzących swoje korzenie z matematyki i mających mocne oparcie w nauce. Warto zastanowić się jednak, z czego na co dzień korzystają osoby zajmujące się weryfikacją jakości danych. Przygotowując niniejszy artykuł, rozmawialiśmy z kilkoma z nich na ten temat.

W ramach badawczej praktyki stosowanej w komercyjnych instytucjach badawczych⁸⁷ stosuje się metody najszybsze, wskazane w części poświęconej metodom niestatystycznym. Taką procedurą może być – i najczęściej jest – kontrola logiczna odbywająca się zgodnie z określonym schematem, różnym dla różnych firm. Na przykład w firmie respondent A jako obserwacje odstające traktuje się wyłącznie górny 1% wszystkich wartości na danej zmiennej, gdzie standardem jest sformułowanie $\pm n\%$ oznaczające zarówno górny, jak i dolny odsetek. Również kontrola logiczna opisana w początkowej części artykułu prowadzona jest w sposób uproszczony. Zwykle już po zakończeniu realizacji danego badania sprawdza się wyłącznie rozkłady odpowiedzi oraz wybrane zależności pomiędzy określonymi zmiennymi.

⁸⁷ Respondent A: mężczyzna, lat 41, wykształcenie wyższe, kierownik zespołu analitycznego w komercyjnym instytucie badań rynkowych od 7 lat, całkowity staż pracy na stanowiskach związanych z kontrolą i analizą danych: 13 lat. Respondent nie wyraził zgody na udostępnienie danych osobowych w postaci imienia i nazwiska oraz nazwy firmy.

Nieco inaczej kontrola logiczna przebiega w przypadku dużych badań naukowych⁸⁸, takich jak na przykład Polski Generalny Sondaż Społeczny czy Europejski Sondaż Społeczny. Przede wszystkim cykl badawczy takich badań jest wydłużony. Podczas gdy cały proces badania prowadzonego przez komercyjny instytut zwykle trwa około dwóch miesięcy – od przygotowania niezbędnych narzędzi do zaprezentowania wyników, to w przypadku badań naukowych, szczególnie wymienionych wyżej, mowa najczęściej o co najmniej dwóch latach. Kontrola takich badań odbywa się jeszcze często z wykorzystaniem papierowych kwestionariuszy, na których doskonale widać wszystkie próby zafałszowania⁸⁹ dokonywane przez ankietera. Taka kontrola jest jednym z pierwszych kroków. Drugim jest kontrola logiczna – na przykład przejść pomiędzy pytaniami, kompletności wypełnienia, powtarzalności odpowiedzi⁹⁰, jakości pytań otwartych. Dopiero później socjologowie sięgają po metody i testy statystyczne. Respondent B poleca jak najczęstsze korzystanie właśnie z tych ostatnich. Szczególnie w sytuacji, gdy badacz nie ma zbyt dużego doświadczenia, a uzyskanie wsparcia takiego specjalisty może być problematyczne. Ponadto respondent B zaleca zaplanowanie badania w taki sposób, żeby pomiędzy poszczególnymi czynnościami pozostały wolne okienka czasowe.

Zupełnie inne podejście reprezentuje respondent C⁹¹. Instytucja, w której on pracuje, podlega jednemu z resortów siłowych, a duża część danych pozyskiwana jest w wyniku pracy operacyjnej. Pierwszym krokiem, który wykonuje jego zespół, jest próba zdobycia tych samych danych z innego, niepowiązanego i zweryfikowanego źródła. Jeśli to zawiedzie, następuje sięganie po modele niepoprawnie nazywane w jego

⁸⁸ Respondent B: kobieta, lat 62, samodzielny pracownik naukowy, kierownik instytutu, uczelnia państwowa, jako pracownik naukowy związana z uczelniami od ponad 35 lat. Respondentka nie wyraziła zgody na udostępnienie danych osobowych w postaci imienia i nazwiska oraz nazwy uczelni.

⁸⁹ W przypadku osób, które – podobnie jak autor (K.K. Kuźma) – zajmują się kontrolą danych od kilkunastu lat, jest to możliwe do wychwycenia również w elektronicznej bazie danych.

⁹⁰ Na odbywającym się w Szczecinie w 2013 r. Zjeździe Socjologicznym jeden z profesorów zajmujących się na co dzień metodologią i metodami nauk społecznych mówił o tym, że ankieter jest w stanie jedną ankietę jeszcze dobrze sfałszować, przy dwóch zaczyna się robić trudno, a przy trzech lub więcej można to łatwo wychwycić.

⁹¹ Respondent C: mężczyzna, lat 39, doktor, kapitan, kierownik zespołu odpowiedzialnego za weryfikację danych operacyjnych. Ze wspomnianym resortem związany od blisko 20 lat. Respondent nie wyraził zgody na udostępnienie danych osobowych w postaci imienia i nazwiska oraz nazwy organizacji.

organizacji ekonometrycznymi. Model ekonometryczny jest bowiem modelem matematyczno-statystycznym, a sama ekonometria jest bardzo często mylona ze statystyką. W rzeczywistości są to właśnie modele statystyczne, w dużej mierze korzystające z wnioskowania indukcyjnego, głównie kanonów Milla⁹².

W podstawowej, najczęściej wykorzystywanej wersji postępowania następuje zebranie wszystkich (dostępnych) możliwych przyczyn i/lub skutków badanego zjawiska. Po ich zebraniu, korzystając z kanonów Milla (kanon jedynej zgodności⁹³, kanon jedynej różnicy⁹⁴, kanon zmian współtowarzyszących⁹⁵, kanon połączonej metody zgodności i kanon różnicy reszt), przeprowadza się dalszą analizę, zgodnie z algorytmem kanonów Milla, co pozwala z dużym prawdopodobieństwem⁹⁶ przewidzieć zależności pomiędzy przyczynami/skutkami, co dalej przekłada się na jakość pozyskanych w sposób operacyjny danych.

Podsumowanie

Zarówno autorzy, jak i eksperci, z którymi rozmawiano podczas przygotowania tego artykułu, solidarnie uważają, że omówione metody mają szeroki zakres zastosowania. Najwygodniejszymi i najprostszymi w użyciu będą omawiane w artykule metody statystyczne jako – z jednej strony – dające gwarancję bezpieczeństwa (poprzez swoje osadzenie w matematyce są one trudniejsze do podważenia), a z drugiej – dające względnie jednoznaczny wynik.

Jak pokazuje przykład respondenta C, metody statystyczne – jak choćby omówione kanony Milla – można wykorzystać do oceny jakości danych zebranych w toku pracy operacyjnej. W podobny sposób możemy wykorzystać rozkłady Gaussa i Poissona. Jeśli nasze dane – i w tym miejscu nie jest istotny sposób ich pozyskania – spełniają którykolwiek z nich, to prawdopodobieństwo tego, że są one dobre (poprawne) rośnie skokowo.

⁹² Kanony zostały sformułowane przez Johna Stuarta Milla w 1843 r. jako podstawowe – podówczas – schematy wnioskowania indukcyjnego, to jest takiego, którego celem jest wnioskowanie o prawdziwości na podstawie posiadanych przesłanek. Indukcja eliminacyjna Milla pozwala na wyszukiwanie związków przyczynowych pomiędzy zjawiskami.

⁹³ Dotyczy związków pomiędzy przyczyną a skutkiem zjawiska.

⁹⁴ Występuje, gdy możemy wskazać warunki niezbędne do zaistnienia określonej sytuacji.

⁹⁵ Możemy zastosować wówczas, kiedy możemy obserwować zmiany w natężeniu badanego zjawiska w zależności od różnych sytuacji towarzyszących.

⁹⁶ Zgodnie z informacjami od respondenta C – sprawdzalność przewidywań z wykorzystaniem kanonów Milla sięga około 85%.

Analogicznie jest w przypadku omawianych testów statystycznych. Jednym z przykładów omawianych w artykule jest liczba podpisanych umów raportowana przez pracownika. Test pozwoli wykryć, czy liczba umów nie jest sztucznie zawyżona. Może on także posłużyć do sprawdzenia wielkości wydatków czy – na przykład – weryfikacji liczby kontaktów osoby, której badaniem się zajmujemy.

Podsumowując, każda z metod ma swoje zastosowanie. A raczej wiele możliwych zastosowań. Staraliśmy się w artykule przedstawić w miarę szeroki katalog możliwości. Zdajemy sobie jednak sprawę, że temat nie został wyczerpany i pozostawia szerokie pole możliwości jego kontynuacji w kolejnych publikacjach.

STRESZCZENIE

Artykuł porusza problem jakości danych otrzymywanych w procesie badań ilościowych z wykorzystaniem kwestionariuszy. Autorzy postanowili poruszyć ten temat, ponieważ w polskiej politologii – w opozycji do socjologii – zajmuje on stosunkowo mało miejsca. Tekst powstał głównie na bazie przeglądu literatury poświęconej poszczególnym poruszonym w nim problemom i stanowi jej syntezę. W trakcie prac odkryto, że dzięki stosunkowo prostym narzędziom, a także wykorzystaniu programów (na przykład Excel czy SPSS) można w łatwy i precyzyjny sposób sprawdzić jakość zbioru danych.

Konrad Gahuszko, Joanna Lewczuk, Konrad Krystian Kuźma

VALIDATE? VERIFY? DO NOT MOVE? ABOUT NON-STATISTICAL, SCHOLASTIC AND QUANTITATIVE METHODS OF QUANTITATIVE DATA STORYTELLING

The article deals with the quality of data obtained in the quantitative research process. The authors decided to raise the subject, because in the Polish political science – in the opposition to sociology – it's not well described. The text was made mainly on the basis of a literature review devoted to particular parts of the article and is its synthesis. During the work it was discovered that thanks to relatively simple tools, as well as the use of some programs the quality of the data set can be checked in a simple and precise way.

KEY WORDS: *data quality assessment, statistical methods, logical control, statistical tests, normal distribution*

Bibliografia

- Barnes S.A., *Identifying Fraud Can Be as Easy as 1, 2, 3: Applying Benford's Law to Forensic Analyses and Investigations*, <https://files.constantcontact.com/85e69272601/6ec4247f-14c0-49cf-92a6-674a5a4c37c2.pdf> (dostęp: 19.01.2019).
- Benford F., *The Law of Anomalous Numbers*, „Proceedings of the American Philosophical Society” 1938, nr 78(4).
- Blattman C. i in., *Measuring the Measurement Error. A Method to Qualitatively Validate Survey Data*, <https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31847/Measuring%20the%20measurement%20error.pdf?sequence=2&isAllowed=y> (dostęp: 20.02.2019).
- Bławat F., *Podstawy analizy ekonomicznej. Teorie, przykłady, zadania*, Warszawa 2011.
- Brzeziński J., *Metodologia badań psychologicznych*, Warszawa 2005.
- Carballo M., Hjelm U. (red.), *Public Opinion Polling in a Globalized World*, Berlin 2008.
- Cartledge P., Garnsey P., Gruen E.S. (red.), *Hellenistic Constructs: Essays in Culture, History, and Historiography*, Berkeley – Los Angeles – Londyn 1997.
- Domański H., Dukaczewska A., *Stabilność odpowiedzi w badaniach socjologicznych*, „ASK. Research and Methods” 1996, nr 1.
- Gostkowski Z., *Analiza „efektu panelowego” w badaniach wyborczych w Łodzi w 1961 r.*, [w:] tegoż (red.), *Analizy i próby technik badawczych w socjologii*, t. 1, Wrocław – Warszawa – Kraków 1966.
- Jabkowski P., *Reprezentatywność badań reprezentatywnych. Analiza wybranych problemów metodologicznych oraz praktycznych w paradygnacie całkowitego błędu pomiaru*, Poznań 2015.
- Mebane W.R. Jr., Kalinin K., *Comparative Election Fraud Detection*, APSA 2009 Toronto Meeting Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450078 2009 (dostęp: 19.01.2019).
- Mider D., Marcinkowska A., *Analiza danych ilościowych dla politologów. Praktyczne wprowadzenie z wykorzystaniem programu GNU PSPP*, Warszawa 2013.
- Młodinow L., *Matematyka niepewności. Jak przypadki wpływają na nasz los*, przekł. P. Strzelecki, Warszawa 2009.
- Nowak E., *Teoria „agenda setting” a nowe media*, „Studia Medioznawcze” 2016, nr 3(66).
- Pawłowski T., *Logiczne podstawy weryfikacji wewnętrznej badań kwestionariuszowych*, [w:] Z. Gostkowski, J. Lutyński (red.), *Wywiad kwestionariuszowy w świetle badań metodologicznych*, *Analizy i Próby Technik Badawczych w Socjologii*, t. 4, Wrocław 1972.
- Sambridge M., Tkalčić H., Arroucau P., *Benford's Law of First Digits: From Mathematical Curiosity to Change Detector*, „Asia Pacific Mathematics Newsletter” 2011, nr 1.
- Sztabiński F., *Ocena jakości danych w badaniach surveyowych*, Warszawa 2011.
- Taylor J.R., *Wstęp do analizy błęd pomiarowego*, przekł., A. Babiński, R. Bożek, Warszawa 2018.