

Krajowy
Instytut
Mediów



Badanie Założycielskie

Usługi medialne
i infrastruktura do ich odbioru

**Podręcznik Badania
w roku 2023**



Spis treści

1. Podstawowe pojęcia.....	4
Procedura konstruowania próby.....	13
2. Wprowadzenie	14
3. Badana populacja i jednostka badania.....	15
4. Procedura konstruowania próby.....	15
5. Przygotowanie do realizacji Badania Założycielskiego	21
5.1. Kwestionariusze.....	21
5.1.1. Kwestionariusz gospodarstwa domowego.....	21
5.1.2. Kwestionariusz indywidualny	23
5.2. Szkolenie ankietatorów	24
5.3. Badanie pilotażowe	25
6. Realizacja badania w terenie.....	27
6.1. Technika zbierania danych	27
6.2. Sieć ankieterska	28
6.3. Sposób realizacji próby.....	28
6.4. Wsparcie techniczne	28
6.5. Działania zwiększające współczynnik realizacji próby (<i>response rate</i>)	29
7. Kontrola realizacji badania	32
8. Przetwarzanie danych	35
9. Konstrukcja systemu wag - uogólnianie wyników na badaną populację	37
10. Ocena precyzji estymacji - uogólnianie wyników na badaną populację	40
11. Udostępnianie wyników badania	43
12. Szacunek uniwersów	45



P₁ O₁ D₁

S₄ T₁ A₁

W₁ O₁ W₁ E₂

P₁ O₁

J₁ Ę₁ C₁ I₁ A₁



1. Podstawowe pojęcia

Alokacja proporcjonalna próby. Alokacja jednostek z próby tak, aby każda warstwa miała reprezentację proporcjonalną do swego udziału w populacji.

Alokacja próby. Jeden z etapów warstwowego losowania próby, polegający na rozmieszczeniu elementów próby w poszczególnych warstwach.

Badanie Założycielskie (BZ) (*establishment survey*). Badanie ciągłe, populacyjne, bazujące na losowej próbie gospodarstw domowych o dużej liczebności, mające na celu ustalenie struktury socjodemograficznej członków gospodarstw oraz stanu wyposażenia gospodarstw domowych i ich członków w środki do odbioru usług medialnych (treści audio-wizualnych) i treści w internecie, a także identyfikację sposobów i częstości korzystania członków gospodarstw domowych z tych usług. Badanie Założycielskie jest podstawą dla ustalenia liczby odbiorców poszczególnych usług medialnych, tzw. uniwersów, szacowanych na podstawie próby badawczej, ekstrapolowanych na objętą badaniem populację gospodarstw domowych oraz na populację mieszkańców Polski w wieku 4 lata i więcej.

Błąd braku odpowiedzi (*non-response error*). Brak odpowiedzi badanej jednostki pojawia się, kiedy nie zgromadzono dla niej kluczowych danych (jednostka nie udzieliła wywiadu lub nie zostały dla niej zebrane dane dotyczące niektórych, lecz nie wszystkich zmiennych).

Błąd całkowity. Suma błędu losowego i błędu systematycznego, jakimi obarczony jest wynik badania statystycznego.

Błąd losowy. Błąd zmieniający się w sposób przypadkowy zarówno co do wartości, jak i co do znaku przy wielokrotnym powtarzaniu pomiaru danej wielkości w niezmiennych warunkach. Wielkość błędów losowych zależy od liczebności próby, zróżnicowania badanej cechy i od wykorzystanego schematu losowania.

Błąd pokrycia populacji. Należy do systematycznych błędów badania i pojawia się, gdy operat losowania nie zawiera niektórych jednostek z badanej populacji lub zawiera jednostki spoza badanej populacji.



Błąd systematyczny (niełosowy). Błąd, który przy wielokrotnym pomiarze danej wielkości w niezmiennych warunkach, pozostaje stały co do wartości i co do znaku, albo zmienia się według znanej zależności. Błędy systematyczne mogą wiązać się z określonym schematem losowania próby lub też mogą pochodzić z innych źródeł. W pierwszym przypadku, gdy dokona się właściwego wyboru schematu losowania próby, błędy są niewielkie i nie wpływają istotnie na końcowe wyniki. Błędy systematyczne niezwiązane ze schematem losowania mogą powodować znaczne zniekształcenie wyników badania.

CAPI - wywiad bezpośredni wspomagany komputerowo (*Computer Assisted Personal Interview*). Technika realizacji wywiadów bezpośrednich, w której kwestionariusz wywiadu jest zaprogramowany i wyświetlany w formie interaktywnego skryptu na komputerze/laptopie. Odpowiedzi respondenta są zaznaczane w interaktywnym skrypcie a sposób udzielania odpowiedzi na pytania i przejścia pomiędzy pytaniami są zaprogramowane i kontrolowane przez skrypt ankiety.

Członek gospodarstwa domowego. Osoba zamieszkała pod danym adresem i utrzymująca się wspólnie z innymi osobami należącymi do jego gospodarstwa domowego.

Cechy warstwujące. Cechy dodatkowe, na podstawie których następuje podział populacji na warstwy (losowanie warstwowe). Cechy warstwujące są powiązane z badaną zmienną - mogą to być cechy demograficzne, społeczne, ekonomiczne, itd.

Dane nieważone. Dane zebrane bezpośrednio w badaniu (dane surowe), nieuwzględniające korekt związanych z doбором próby i jej realizacją w terenie.

Dane ważne. Dane uwzględniające korekty wyników z próby, prowadzące do uzyskania istotnej poprawy zgodności rozkładów wybranych cech badanych jednostek w próbie z ich rozkładami w populacji generalnej.

Efekt ankieterski. Zafałszowanie odpowiedzi udzielanych przez respondenta, spowodowane jego reakcją na sposób zachowania lub wygląd ankietera (sugerowanie lub ocenianie odpowiedzi, indywidualne modyfikacje procedury badania, niekulturalne słownictwo, niegrzeczne zachowanie, wiek lub płeć ankietera, jego ubiór, niechlujny wygląd). Czasami takie zafałszowanie powstaje również w wyniku błędnych przeświadczeń respondenta na temat oczekiwań ankietera wobec jego osoby.



Estymacja parametryczna. Szacowanie wartości parametrów z populacji generalnej na podstawie obserwacji uzyskanych w próbie losowej. Obejmuje ona estymację punktową i estymację przedziałową.

Estymacja punktowa. Wynikiem estymacji punktowej jest konkretna wartość (liczba), otrzymana na podstawie wyników próby, stanowiąca ocenę wartości parametru.

Estymacja przedziałowa. Wynikiem estymacji przedziałowej jest przedział liczbowy, w którym z określonym prawdopodobieństwem zawiera się wartości parametru z populacji generalnej.

Estymator. Funkcja, na podstawie której w oparciu o wyniki próby szacujemy wartość danego parametru (cechy) w populacji generalnej. Przykładowo, estymatorem wartości średniej arytmetycznej w populacji generalnej (wartości oczekiwanej) jest średnia arytmetyczna z próby.

Gospodarstwo domowe. Zespół osób zamieszkujących razem i wspólnie utrzymujących się (wg definicji z NSP 2011). Osoby samotne utrzymujące się samodzielnie tworzą jednoosobowe gospodarstwa domowe.

Grupa społeczno-ekonomiczna gospodarstwa domowego. Inaczej źródło utrzymania gospodarstwa domowego określane przez wyłączne lub główne (przeważające) źródło uzyskania dochodów (lub innych przychodów) członków gospodarstwa domowego.

Jednostka badania. Element populacji generalnej podlegający badaniu (np. gospodarstwo domowe, osoba).

Jednostka losowania. Jednostka wylosowana na określonym stopniu losowania w ramach danego schematu losowania. Jednostka losowania może być pojedynczą jednostką badania lub może składać się z pewnej liczby jednostek badania.

Klasyfikacja Jednostek Terytorialnych do Celów Statystycznych (NUTS). Uporządkowany wykaz nazw jednostek terytorialnych i przypisanych im symboli terytorialnych występujących na poszczególnych poziomach podziału terytorialnego kraju, wykorzystywany w procesie zbierania opracowywania i rozpowszechniania porównywalnych danych na obszarze Unii Europejskiej dla określonych statystyk regionalnych. Klasyfikacja NUTS ma charakter hierarchiczny – dzieli każde państwo członkowskie Unii na jednostki terytorialne poziomu NUTS



1, z których każda dzieli się na jednostki terytorialne poziomu NUTS 2, a te z kolei dzielą się na jednostki terytorialne poziomu NUTS 3, przy czym jedna jednostka terytorialna może reprezentować kilka poziomów NUTS. Założeniem klasyfikacji jest to, aby jednostki każdego z poziomów ze wszystkich państw Unii Europejskiej były zbliżone do siebie pod względem liczby ludności. Od 1 stycznia 2018 roku w Polsce funkcjonuje 97 jednostek NUTS:

NUTS 1 – makroregiony (grupujące województwa) – 7 jednostek

NUTS 2 – regiony (województwa lub ich części) – 17 jednostek

NUTS 3 – podregiony (grupujące powiaty w ramach województw) – 73 jednostki

Kwestionariusz wywiadu. Narzędzie badawcze zawierające uporządkowany zbiór pytań, które ankieter zadaje respondentowi, listę odpowiedzi (pytania zamknięte) lub umożliwiające zapis swobodnych odpowiedzi respondenta (pytania otwarte) oraz instrukcje dla osoby prowadzącej wywiad/respondenta.

Liczebność próby. Liczba jednostek wylosowana do próby, podlegających bezpośrednio badaniu.

Lider gospodarstwa domowego. W Badaniu Założycielskim realizowanym przez Krajowy Instytut Mediów, osoba, która ukończyła 18 rok życia, wyznaczona przez członków gospodarstwa domowego jako ta, która posiada największą wiedzę na temat osób w gospodarstwie domowym oraz wyposażenia gospodarstwa domowego i jego członków w urządzenia do odbioru poszczególnych mediów elektronicznych, źródeł i dostawców sygnału elektronicznych usług mediowych.

Losowanie bez zwracania. Losowanie proste indywidualne, w którym jednostka wylosowana do badania nie bierze już udziału w dalszym losowaniu.

Losowanie jednostek do próby. Technika pobierania jednostek do próby realizująca założenia losowego wyboru próby.

Losowanie metodą Hartleya-Rao. Systematyczny dobór jednostek do próby, po ich uprzednim uporządkowaniu.

Losowanie proste. Losowanie z populacji generalnej z jednakowymi prawdopodobieństwami indywidualnych jednostek badania w sposób bezpośredni, gdzie wszystkie jednostki losowania są jednocześnie jednostkami badania.



Losowanie systematyczne. Polega na wybraniu do próby co n -tej jednostki znajdującej się w operacie losowania. Kluczową cechą tego schematu losowania jest fakt, iż jeżeli jednostki populacji zostaną uporządkowane według cech mających bezpośredni związek z badaną cechą, to schemat ten zwiększa reprezentatywność próby.

Losowanie warstwowe (dwustopniowe). Dwustopniowa procedura, zgodnie z którą populację dzieli się na subpopulacje (podpopulacje), zwane warstwami, a następnie z poszczególnych subpopulacji niezależnie pobiera się w sposób losowy elementy do próby. Celem podziału badanej populacji na subpopulacje, czyli inaczej celem warstwowania, jest wyodrębnienie w zróżnicowanej zbiorowości możliwe jednorodnych grup jednostek. Dzięki temu zabiegowi każda z tych grup będzie miała odpowiednią reprezentację w próbie.

Losowanie ze zwracaniem. Losowanie proste indywidualne, w którym nadajemy każdej jednostce statystycznej jednakowe prawdopodobieństwo dostania się do próby. Jednostka raz wylosowana do badania bierze udział w dalszym losowaniu.

Losowy wybór próby. Technika pobierania próby z badanej populacji generalnej, która spełnia dwa następujące warunki: 1) każda jednostka populacji ma dodatnie i znane prawdopodobieństwo dostania się do próby, 2) dla każdego zespołu jednostek populacji można ustalić prawdopodobieństwo tego, że w całości znajdzie się on w próbie. Pierwszy warunek odnosi się do sytuacji wyboru indywidualnych jednostek do badania (osób, gospodarstw, przedsiębiorstw, rzeczy), drugi warunek dotyczy sytuacji, gdy losuje się całe zespoły jednostek (np. grupy studenckie, obwody spisowe). Należy podkreślić, że tylko próba losowa pozwala ocenić precyzję oszacowania wartości parametrów w populacji generalnej.

Metoda badań. Planowy i powtarzalny sposób zbierania danych empirycznych.

Narodowy Spis Powszechny (właściwie: Narodowy Spis Powszechny Ludności i Mieszkań). Podstawowe badanie pełne i źródło danych, mające na celu zebranie informacji o stanie i strukturze ludności według ustalonych cech demograficznych i społeczno-zawodowych, w określonym momencie i na określonym terytorium. Powszechny spis ludności powinien spełniać trzy warunki: powszechności (obejmować całą ludność), jednoczesności (musi być przeprowadzony w określonym czasie), imienności (każda osoba ma zostać spisana z imienia i nazwiska). Dodatkowo w Polsce stosuje się zasady regularności (spisy odbywają się wg zaleceń ONZ co 10 lat) oraz bezpośredniości



(odpowiedzi na pytania spisowe udziela bezpośrednio osoba spisywana, w wyjątkowych sytuacjach jej najbliżsi domownicy). Dostarcza on m.in. unikalnych informacji o gospodarstwach domowych i ich członkach, wykorzystywanych do prognoz demograficznych. W chwili obecnej dostępne są nadal pełne dane ze spisu przeprowadzonego w 2011 r. Dane ze Spisu Powszechnego 2021 są dostępne wybiórczo, co uniemożliwia pełne ich zastosowanie (dotyczy to w szczególności estymowania populacji gospodarstw i projekcji na przyszłość).

Nielosowy wybór próby. Technika doboru próby z populacji generalnej, która nie spełnia choć jednego z dwóch warunków określonych w definicji losowego wyboru próby. Najpopularniejszymi technikami nielosowego wyboru próby są: wybór przypadkowy, wybór celowy i wybór kwotowy.

Obwód spisowy. Jednostka przestrzenna wyodrębniona dla spisów powszechnych i innych badań statystycznych według liczby mieszkań i mieszkańców. Pod kątem sprawnego przeprowadzenia czynności spisowych przez jednego rachmistrza spisowego ustalono kryterium wielkości obwodu spisowego wynoszące nie więcej niż 500 osób i 200 mieszkań. Obwód spisowy wchodzi w skład rejonu statystycznego; maksymalna liczba obwodów w rejonie nie przekracza 9.

Operat losowania próby. Wykaz jednostek badania lub ich zespołów (zespoły te muszą być rozłączne) zwanych jednostkami losowania wraz z ich identyfikacją, spełniający warunek, że każdej jednostce losowania odpowiada tylko jeden numer i jednocześnie każdemu numerowi odpowiada tylko jedna jednostka losowania. Jeżeli wszystkie jednostki losowania są jednocześnie jednostkami badania, to mówimy wtedy o losowaniu indywidualnym. Jeżeli warunek ten nie jest spełniony, to losowanie nazywamy zespołowym. Operatem losowania próby w Badaniu Założycielskim jest TERYT.

Plan losowania próby. Mechanizm losowania realizujący sposób wyboru elementów populacji do próby.

Podpopulacja (subpopulacja). Wybrany w określony sposób podzbiór populacji, spełniający określone kryteria lub kryterium, stanowiący w dalszej części badania nową populację.



Populacja generalna (zbiorowość generalna). Zbiór (agregat) dowolnych elementów (osób, gospodarstw domowych, przedmiotów, zdarzeń) powiązanych ze sobą logicznie (posiadających wspólną cechę) ale nieidentycznych z punktu widzenia badanej cechy (lub cech). Populacja generalna obejmuje wszystkie elementy będące przedmiotem badania, w odniesieniu do których formułuje się wnioski ogólne.

Próba. Podzbiór badanej populacji (obejmujący część jej elementów), bez względu na sposób wyboru jego elementów, będący przedmiotem badania.

Próba reprezentatywna. Próba, której struktura ze względu na badane cechy (zmienne) jest zbliżona do struktury populacji, z której pochodzi. Reprezentatywność próby można uzyskać stosując zarówno losowe (probabilistyczne) jak i nielosowe (nieprobabilistyczne) techniki wyboru próby. Należy przy tym zaznaczyć, iż większą szansę na reprezentatywność próby daje zastosowanie technik losowego jej doboru.

Pytanie zamknięte. Pytanie, w którym oczekuje się od respondenta jedynie wskazania jednego lub kilku spośród podanych wariantów odpowiedzi.

Pytanie otwarte. Pytanie, na które respondent może udzielić swobodnej odpowiedzi, nie warunkowanej sugerowanymi wariantami odpowiedzi.

Respondent. Uczestnik badania, czyli osoba, która odpowiada na pytania ankietera, moderatora lub samodzielnie wypełnia ankietę pocztową, internetową itp. Respondent nie jest koniecznie jednostką badania – może to być przykładowo osoba udzielająca odpowiedzi w imieniu gospodarstwa domowego lub dziecka pozostającego pod jego opieką.

Rejon statystyczny. Jednostka przestrzenna agregacji danych statystycznych, złożona z kilku, nie więcej niż dziewięciu, obwodów spisowych. Kryterium wielkości rejonu statystycznego jest liczba mieszkań i ludności wynosząca nie więcej niż 2700 osób i 999 mieszkań.

Schemat losowania próby. Proces wyboru jednostek statystycznych z populacji generalnej ze z góry określonym zbiorem prawdopodobieństw wyboru dla poszczególnych jednostek w każdym losowaniu.



Standardowy błąd szacunku. Miara rozrzutu wartości oszacowań parametrów na podstawie próby wokół wartości tego parametru z populacji generalnej. Im mniejsza jego wartość tym dokładniejszy pomiar.

System Zarządzania Badaniami KIM. Narzędzie informatyczne Krajowego Instytutu Mediów wspomagające realizację Badania Założycielskiego.

Technika badań. Konkretny, w ramach danej metody lub jej wariantu, sposób gromadzenia danych empirycznych

TERYT. Krajowy Rejestr Urzędowy Podziału Terytorialnego Kraju, prowadzony przez GUS. TERYT zawiera:

- identyfikatory i nazwy jednostek podziału terytorialnego (województwo, powiat, gmina),
- identyfikatory nazw miejscowości,
- identyfikatory rejonów statystycznych i obwodów spisowych,
- identyfikację adresową ulic, nieruchomości, budynków i mieszkań.

Uniwersum w Badaniu Założycielskim. Określona liczebnie grupa odbiorców poszczególnych usług medialnych, szacowana na podstawie próby badawczej, ekstrapolowanych na objętą badaniem populację gospodarstw domowych oraz na populację mieszkańców Polski w wieku 4 lata i więcej.

Warstwa. Podpopulacja (zwana też subpopulacją) badanej populacji. Wybrany w określony sposób podzbiór populacji stanowiący w dalszej części badania nową populację (losowanie warstwowe).

Wagi. Współczynniki liczbowe przypisywane w toku analizy danych poszczególnym badanyom jednostkom w celu dopasowania struktury próby do struktury populacji ze względu na wybrane cechy badanych jednostek. Różnice pomiędzy tymi strukturami mogą powstawać zarówno na etapie dobierania próby, jak i jej realizacji w terenie.

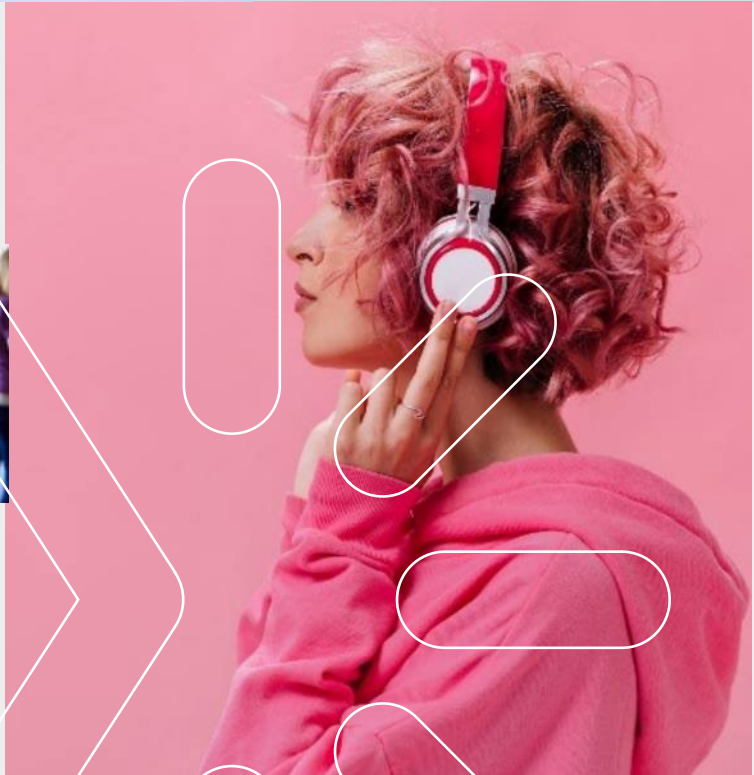
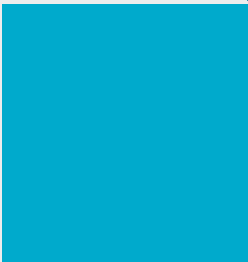
Ważenie wyników badania. Korygowanie wyników z badanej próby w taki sposób, aby były w jak największym stopniu reprezentatywne dla populacji, z której została pobrana próba.

Ważenie zintegrowane. Proces ważenia wyników z próby uwzględniający jednocześnie dane (z próby i z populacji) dotyczące gospodarstw domowych i ich członków.



Wiązka realizacyjna. Grupa adresów skupionych przestrzennie, przeznaczona najczęściej do zrealizowania przez ankietera w trakcie jednego dnia pracy (zadanie badawcze). Wywiady takie są zwykle zlokalizowane w niewielkiej odległości – w BZ są to wywiady realizowane w ramach jednego rejonu statystycznego.

Wskaźnik odpowiedzi (response rate). Poziom realizacji próby mierzony stosunkiem liczby zbadanych jednostek do liczebności wylosowanej próby, wyrażany w procentach. w Badaniu Założycielskim stosunek liczby gospodarstw domowych, w których zrealizowano komplet wywiadów do liczby odwiedzonych adresów pomniejszonej o jednostki nienależące do badanej populacji.



Wprowadzenie

Badana populacja
i jednostka badania

Procedura konstruowania próby

2. Wprowadzenie

Krajowy Instytut Mediów (KIM) rozpoczął w 2021 r. realizację Badania Założycielskiego (BZ), obejmującego konsumpcję usług medialnych w Polsce. Podstawowym celem BZ jest uzyskanie kompleksowego obrazu konsumpcji usług medialnych przez gospodarstwa domowe i ich członków w zakresie:

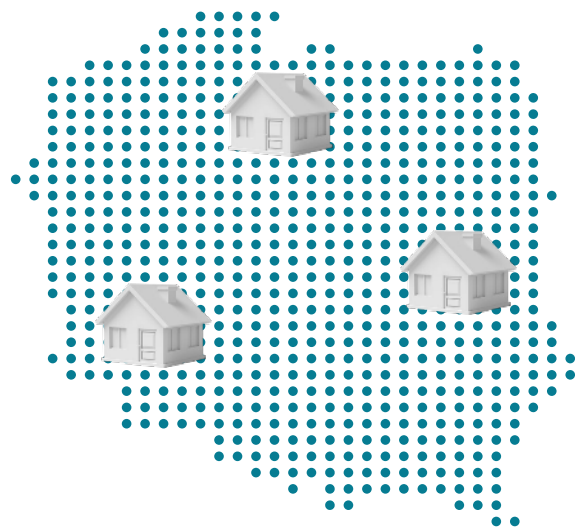
- wyposażenia gospodarstw domowych i ich członków w urządzenia do odbioru poszczególnych usług medialnych,
- sposobu i częstości konsumpcji poszczególnych usług medialnych i współkorzystania z usług przez członków gospodarstw domowych w zależności od ich charakterystyk demograficznych i społeczno-ekonomicznych (m. in. płeć, wiek, poziom wykształcenia, aktywność zawodowa, miejsce zamieszkania, a także wielkość, struktura i poziom dochodów ich gospodarstwa domowego),
- źródeł i dostawców sygnału i infrastruktury do odbioru usług medialnych,
- powodów niekorzystania z usług medialnych przez członków gospodarstw domowych w zależności od ich charakterystyk demograficznych i społeczno-ekonomicznych.

Wyniki Badania Założycielskiego zostaną wykorzystane do:

- oszacowania uniwersów usług medialnych, tj. struktury gospodarstw domowych i ich członków według cech opisujących konsumpcję poszczególnych usług medialnych, które to dane nie są dostępne w zbiorach danych statystyki publicznej,
- tworzenia profili odbiorców poszczególnych usług medialnych,
- wyznaczenia zmiennych kontrolnych dla innych badań dotyczących usług medialnych,
- oceny gotowości użytkowników usług medialnych do wdrażania nowości technicznych (m. in. wdrożenia systemu kodowania telewizji naziemnej DVB-T2, HEVC, nadawania sygnału radiowego w technologii DAB+).

3. Badana populacja i jednostka badania

Populacją objętą Badaniem Założycielskim są wszystkie gospodarstwa domowe zamieszkujące w Polsce w trakcie prowadzenia badania. Nie podlegają badaniu gospodarstwa domowe zamieszkałe w obiektach zbiorowego zakwaterowania (internat, hotel robotniczy, dom rencisty, klasztor itp.) z wyjątkiem gospodarstw domowych pracowników zamieszkałych w tych obiektach z tytułu wykonywanej pracy (np. kierownik hotelu, dozorca). Badaniem objęte są także gospodarstwa domowe obcokrajowców, z wyłączeniem sytuacji, gdy niemożliwa jest realizacja badania w języku polskim.



Jak wskazują wyniki wielu badań znacząca część konsumpcji usług medialnych, szczególnie telewizji i radia dalej odbywa się za pośrednictwem dostępnego w gospodarstwie telewizora i radia. Jednocześnie konsumenci coraz częściej korzystają z usług medialnych wykorzystując internet i urządzenia mobilne, czyniąc konsumpcję coraz bardziej indywidualną. Aby uwzględnić zarówno aktualnie istniejący sposób konsumpcji usług medialnych (używanych źródeł i odbiorników), jak i zachodzące w nim zmiany, jednostką badania w Badaniu Założycielskim jest gospodarstwo domowe i wszyscy jego członkowie, którzy w momencie badania ukończyli 4 lata.

4. Procedura konstruowania próby

Badanie Założycielskie było realizowane przez cały 2023 rok. w trakcie roku zebrano informacje o 15 015 gospodarstwach domowych. Zebrano również 32 735 wywiadów indywidualnych, zrealizowanych z członkami tych gospodarstw w wieku 4 lata i więcej.

Badanie Założycielskie jest prowadzone metodą reprezentacyjną, w której elementy populacji generalnej zostały dobrane do próby za pomocą mechanizmu losowego co oznacza, że każda jednostka losowania ma znane prawdopodobieństwo wyboru do próby. Z teoretycznego i praktycznego punktu widzenia metoda reprezentacyjna jest najbardziej prawidłową formą badania częściowego. Tylko na bazie próby losowej można, na podstawie jej wyników, oszacować właściwości (cechy) populacji generalnej, czyli uogólnić uzyskane na podstawie próby wyniki na całą populację generalną (estymacja punktowa) wraz z błędami szacunku tych parametrów (estymacja przedziałowa). Wszystkie te elementy wnioskowania odwołują się w sposób bezpośredni do zasad rachunku prawdopodobieństwa, które nie mają zastosowania, gdy nie operujemy próbą losową.

Podstawę losowania jednostek do próby (operat losowania) w Badaniu Założycielskim stanowił system TERYT (Urzędowy Rejestr Podziału Terytorialnego Kraju), prowadzony przez Główny Urząd Statystyczny, który jako jedyny dostępny operat losowania mógł być zastosowany do losowego doboru próby przy założeniu, że podstawową jednostką badania jest gospodarstwo domowe. Obejmuje on adresy wszystkich mieszkań i domów w Polsce wraz z informacją o aktualnym wykorzystaniu lokalu. Pozwala to na etapie doboru próby wyłączyć z niej mieszkania i domy wykorzystywane do innych celów niż mieszkaniowe (np. siedziby firm) oraz czasowo lub trwale niezamieszkałe (w trakcie remontów, przeznaczone do rozbiórki itp.). Do próby gospodarstw domowych, na poziomie spisu, wchodzi wszystkie gospodarstwa domowe zamieszkujące wylosowane do próby mieszkania. Identyfikacja gospodarstw w wylosowanym mieszkaniu odbywa się poprzez zidentyfikowanie przez ankietera na wstępie wywiadu wśród osób zamieszkujących pod wskazanym adresem, grup osób wspólnie utrzymujących się (gospodarstw domowych) i wylosowanie do badania jednego gospodarstwa domowego, z którym oprócz spisu prowadzony jest wywiad. Na bazie próby gospodarstw domowych została utworzona próba osób, członków badanych gospodarstw domowych w wieku 4 lat i więcej w momencie realizacji badania.

W Badaniu Założycielskim zaplanowano zrealizowanie w ciągu roku próby minimum 15 000 adresów wylosowanych z Rejestru TERYT. Dodatkowo przy losowaniu próby dokonano warstwowania z uwzględnieniem 73 podregionów (NUTS-3), co gwarantuje większe rozproszenie próby w terenie i bardziej równomierne. Co do zasady, próby miesięczne



i kumulowane w dłuższych okresach zachowują strukturę populacji z uwzględnieniem warstw określonych przez województwa i klasy wielkości miejscowości.

Próba do Badania Założycielskiego została wylosowana przy zastosowaniu schematu warstwowego dwustopniowego, proporcjonalnego metodą Hartleya-Rao. Losowanie dwustopniowe polega najpierw na losowaniu, według wybranego schematu losowania, jednostek pierwszego stopnia (jps), a następnie w wylosowanych jps-ach dokonujemy losowania, także według wybranego schematu losowania, jednostek drugiego stopnia (jds). Losowanie warstwowe jest efektywniejsze od losowania prostego, gdyż pozwala na ustalenie liczebności losowanych jednostek w poszczególnych warstwach, tak by ich proporcje odzwierciedlały proporcje populacyjne a nie kształtowały się losowo, eliminując w ten sposób jedno ze źródeł błędów losowania prostego. Pozwala ono także na ustalenie w próbie liczebności jednostek z poszczególnych warstw populacji na poziomie zapewniającym żądaną precyzję oszacowań interesujących nas charakterystyk (cech) każdej warstwy. Natomiast wielostopniowość losowania umożliwia zastosowanie różnych schematów losowania na poszczególnych stopniach losowania.

Podstawową warstwę stanowiła gmina. w miastach powyżej 500 tys. mieszkańców (Krakowie, Łodzi, Poznaniu, Wrocławiu oraz Warszawie) losowanie odbywało się na poziomie dzielnic. Oznacza to utworzenie 991 warstw miejskich (w tym 37 dzielnicowych) oraz 73 warstw wiejskich zdefiniowanych przez podregiony NUTS-3. Liczba wylosowanych adresów była proporcjonalna do liczby adresów w danej warstwie.

Jednostkami losowania pierwszego stopnia (jps) były rejony statystyczne. W ramach każdej warstwy (miasta, dzielnicy, wsi w podregionie) losowana była taka liczba rejonów statystycznych, jaka wynika z podzielenia liczebności próby w warstwie przez 4 (liczba adresów losowanych następnie w rejonie). Losowanie rejonów statystycznych odbywało się ze zwracaniem, z prawdopodobieństwem proporcjonalnym do liczby adresów w rejonie statystycznym.

Jednostkami losowania drugiego stopnia (jds) w wylosowanych na I stopniu rejonach były mieszkania. w ramach wylosowanych rejonów statystycznych zostały wylosowane bez zwracania po 4 adresy (bądź ich krotność w przypadku wielokrotnego wylosowania tego samego rejonu).



W tabeli 1 podano liczbę adresów, które zostały wylosowane do próby w poszczególnych klasach wielkości miejscowości.

Tabela 1. Struktura próby adresowej

Klasa wielkości	1	2	3	4	5	6	7	8	9	
Województwo	wieś	< 10 tys.	10-19 tys.	20-49 tys.	50-99 tys.	100-199 tys.	200-499 tys.	500-999 tys.	>1 000 tys.	Razem
dolnośląskie	320	112	96	128	152	48		332		1188
kujawsko-pomorskie	280	52	80	24	72	144	160			812
lubelskie	392	40	60	80	76		164			812
lubuskie	132	48	56	52		120				408
łódzkie	320	36	56	116	128			340		996
małopolskie	544	60	72	104	28	44		396		1248
mazowieckie	664	64	116	228	96	120			860	2148
opolskie	164	48	44	44	24	64				388
podkarpackie	380	48	44	96	76	100				744
podlaskie	156	40	32	40	56		144			468
pomorskie	272	44	28	136	80		336			896
śląskie	368	48	60	216	256	676	256			1880
świętokrzyskie	232	44	28	48	28	92				472
warmińsko-mazurskie	200	44	64	88	24	132				552
wielkopolskie	488	104	108	128	184			272		1284
zachodniopomorskie	192	68	64	112	28	48	192			704
Razem	5104	900	1008	1640	1308	1588	1252	1340	860	15000

Źródło: Obliczenia własne na podstawie NSP 2002, prognozy gospodarstw domowych na lata 2016-2050, oraz rejestru TERYT.

Zastosowany schemat losowania próby zapewnia maksymalne jej rozproszenie terytorialne i równomierne pokrycie całego terytorium Polski, z uwzględnieniem rejonizacji i wielkości miejscowości i jednocześnie optymalizuje koszty realizacyjne. Pierwsze dwa postulaty są realizowane poprzez odpowiednie warstwowanie badanej populacji przy doborze próby, uwzględniające regionalizację do poziomu podregionów– NUTS-3, województw i 9 klas wielkości miejscowości oraz włączenie do próby wszystkich jednostek administracyjnych w kraju. Na etapie warstwowania próby miasta z gmin miejsko-wiejskich zostały włączone do warstw miast odpowiedniej klasy wielkości, a części wiejskie tych gmin do warstw wsi. Postulat optymalizacji kosztów jest spełniony poprzez odpowiednią konstrukcję wiązek realizacyjnych w badaniu (pozwalającej na zmniejszenie średniego czasu potrzebnego

na dotarcie do poszczególnych respondentów i jednocześnie zmniejszenie średniego kosztu dotarcia ankietera do respondenta na kilka wywiadów).

Dla Badania Założycielskiego prowadzonego w 2023 r. podstawą warstwowania próby gospodarstw były dane GUS zebrane w NSP 2002¹ (dostępne dla poszczególnych gmin), dane z Prognozy gospodarstw domowych 2016–2050, jak również zagregowane dane z rejestru TERYT². W odniesieniu do nowoutworzonych miast (nieobjętych NSP 2002 i nieuwzględnionych w prognozach) liczebności próby zostały oszacowane z uwzględnieniem informacji z rejestru TERYT. Pozwala to na optymalne szacowanie aktualnych liczebności populacji gospodarstw domowych w poszczególnych jednostkach administracyjnych – gminach wiejskich / miastach.

W każdym wylosowanym rejonie statystycznym wylosowano dodatkowe adresy stanowiące próbę rezerwową.

¹ Dane z Narodowego Spisu Powszechnego 2011 na poziomie gmin nie są dostępne. Gdy dane będą dostępne na poziomie gmin z Narodowego Spisu Powszechnego 2021, zostaną one wykorzystane do warstwowania próby w kolejnych latach realizacji Badania Założycielskiego.

² Aktualność operatu TERYT na dzień 30.09 poprzedniego roku.

Przygotowanie
do realizacji
Badania
Założycielskiego



5. Przygotowanie do realizacji Badania Założycielskiego

5.1. Kwestionariusze

Badanie Założycielskie prowadzone jest za pomocą zintegrowanych kwestionariuszy gospodarstwa domowego oraz 3 typów kwestionariuszy indywidualnych, których zawartość różni się w zależności od wieku respondentów:

- kwestionariusz dla dzieci, które w momencie badania ukończyły 4 lata i nie ukończyły 13 lat, z tym, że w przypadku wywiadu z najmłodszymi dziećmi na pytania w nim zawarte mogą odpowiadać rodzice / opiekunowie dziecka, w zależności od wieku dziecka
- kwestionariusz dla młodzieży w wieku 13–15 lat,
- kwestionariusz dla osób, które ukończyły 16 lat.

5.1.1. Kwestionariusz gospodarstwa domowego

Podstawowymi elementami kwestionariusza gospodarstwa domowego są:

- część dotycząca realizacji wywiadu,
- część dotycząca wielkości i składu gospodarstwa domowego oraz charakterystyk demograficznych i społeczno-ekonomicznych jego członków (metryczka),
- część merytoryczna zawierająca zestaw pytań odnośnie do wyposażenia gospodarstwa domowego odbiorniki oraz źródła sygnału oraz korzystania przez gospodarstwo domowe z usług medialnych.

W części dotyczącej realizacji wywiadu znajdują się pytania dotyczące:

- zgody gospodarstwa domowego na uczestniczenie w badaniu,
- zgody na wywiad z głową gospodarstwa domowego / inną osobą z gospodarstwa domowego orientującą się w wyposażeniu gospodarstwa domowego w urządzenia do odbioru mediów elektronicznych (liderem technologicznym),
- przyczyn odmów uczestniczenia w badaniu i zgody na wywiad.

Część zawierająca charakterystyki wszystkich osób mieszkających pod adresem (multi HH):

- liczba osób,



- wiek osób
- wykształcenie osób,
- liczba gospodarstw domowych

Część zawierająca charakterystyki gospodarstwa domowego obejmuje następujące ich cechy:

- wielkość gospodarstwa domowego,
- miejsce zamieszkania,
- główne źródło utrzymania,
- przeciętny miesięczny całkowity dochód netto gospodarstwa,
- typ budynku, w którym znajduje się mieszkanie.

Natomiast charakterystyki członków gospodarstwa domowego dotyczą:

- płci,
- wieku,
- stopnia pokrewieństwa z głową gospodarstwa domowego
- poziomu wykształcenia,
- aktywności zawodowej,
- zawodu,
- narodowości,
- niepełnosprawności.

Część merytoryczna kwestionariusza gospodarstwa domowego zawiera pytania dotyczące:

- wyposażenia gospodarstwa domowego w odbiorniki usług medialnych, wykorzystywane w ciągu ostatnich 12 miesięcy,
- wyposażenia gospodarstwa w sprzęt komputerowy i jego charakterystyka
- źródła sygnału telewizyjnego, dostawcy i typu usług;
- dostępu do internetu (m.in. rodzaj dostępu, dostawca,) i typu usług,
- przyczyn niekorzystania z internetu w domu oraz planów uzyskania dostępu do internetu,
- dostępu do serwisów video / platform streamingowych w gospodarstwie domowym,



5.1.2. Kwestionariusz indywidualny

Kwestionariusz dla członków gospodarstwa domowego, którzy ukończyli 16 lat w momencie prowadzenia wywiadu zawiera pytania odnośnie:

- osobistego korzystania z poszczególnych odbiorników usług medialnych, zarówno znajdujących się w domu jak i poza domem, w ciągu ostatnich 12 miesięcy ,
- konsumpcja usług medialnych tj. telewizja, radio, VOD/ streaming
- częstości i czasu w kontekście dnia wczorajszego, poświęconego na osobiste korzystanie z poszczególnych usług medialnych,
- częstości i czasu w kontekście dnia wczorajszego, korzystania z internetu, rodzaju serwisów, stron i aplikacji, akceptacji regulaminu korzystania ze stron internetowych, korzystania z systemu kontroli rodzicielskiej, oceny prawdziwości treści video i ich weryfikacji,
- rodzaju oglądanych programów telewizyjnych i słuchanych programów radiowych, treści w internecie
- czasu poświęconego na czytanie prasy oraz sposobu jej czytania (wydanie papierowe lub e-wydanie), miejsca czytania prasy,
- czytelnictwa książek i ich formy (książki drukowane, ebooki, audiobooki, liczby poczytanych książek w ciągu ostatnich 12 miesięcy),
- charakterystyki stylu życia jednostek obejmującej:
 - ✓ uczestnictwo w kulturze i sposób spędzania czasu (kino, teatr, koncert, filharmonia, muzeum, galeria sztuki, biblioteka, wydarzenia sportowe, festyn, restauracja, czytanie książek),
- zgody na ponowny kontakt,
- przebiegu wywiadu z respondentem (pytania/bloki pytań sprawiające kłopot respondentowi, pytania komentowane przez respondenta).

W przypadku kwestionariuszy dla młodzieży (13-15 lat) nie występuje w nich metryczka oraz blok pytań charakteryzujący ich styl życia. Ponadto z kwestionariusza dla dzieci (4-12 lat) wyłączono pytania dotyczące częstości i czasu poświęconego na korzystanie



z poszczególnych mediów elektronicznych oraz część pytań związanych z korzystaniem z internetu.

5.2. Szkolenie ankierów

Szkolenie ankierów było prowadzone w formie warsztatów, przez kierownika projektu (*project manager*) z KIM. W ramach szkolenia ankierzy zapoznali się z filmami szkoleniowymi (omówienie pełnego pakietu szkoleniowego obejmującego m.in. cel badania, aranżację wywiadu, narzędzia do pracy ankierera, przebieg badania; omówienie zawartości teczek ankierera i manuala ankierera; omówienie zasad bezpieczeństwa dotyczących danych osobowych i informacji przekazywanych przez respondentów oraz bezpieczeństwa zdrowotnego ankierera; możliwości korzystania z internetu/podłączenia do internetu w posiadanych odbiornikach telewizyjnych).



Ankierzy otrzymali narzędzia do pracy ankierera, czyli teczkę ankierera zawierającą:

- list referencyjny,
- ulotkę respondenta,
- ulotkę terminu badania,
- kartę terminu badania,
- kartę ponownej wizyty,
- manual ankierera,
- bon upominkowy,
- klauzulę informacyjną przetwarzania danych osobowych,
- legitymację ankierera.

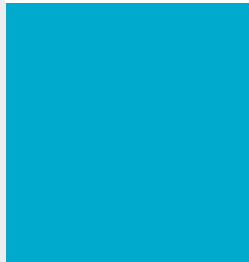
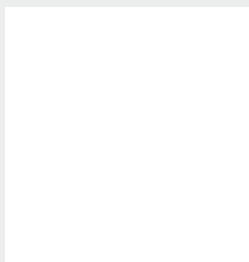
W manualu ankierera obok prezentacji materiałów znajdujących się w tece ankierera znajdowały się wskazówki i podpowiedzi dotyczące sposobu aranżacji wywiadu i jego prowadzenia, przebiegu wywiadu, ułatwiające prowadzenie wywiadu (m.in. „słowa-klucze”

mające zachęcić różne podgrupy respondentów do udziału w badaniu), sposobów radzenia sobie z trudnymi sytuacjami w trakcie wywiadu, sposobu zakończenia wywiadu. Kolejnym elementem szkolenia była obsługa aplikacji internetowej (Systemu Zarządzania Badaniami KIM – w części dotyczącej prowadzenia wywiadu). W ramach tego elementu szkolenia ankieterzy realizowali testową ankietę w tej aplikacji i uczyli się jak uzupełniać rezultaty kontaktu.

5.3. Badanie pilotażowe

Realizacja Badania Założycielskiego jest projektem cyklicznym, realizowanym w sposób ciągły od roku 2022. Pierwsza realizacja w roku 2022 została poprzedzona pilotażem, którym objęto 5 tys. gospodarstw domowych i ich członków w wieku 4 lat i więcej. Próba adresów do badania pilotażowego (plan losowania, schemat losowania, operat losowania) była konstruowana na identycznych zasadach jak próba do Badania Założycielskiego. Zasadniczymi celami badania pilotażowego były:

- uzyskanie wstępnej wiedzy o zbiorowości gospodarstw domowych i ich członków w zakresie wyposażenia gospodarstw domowych w urządzenia do odbioru mediów elektronicznych oraz konsumpcji tych mediów w zależności od charakterystyk demograficznych i społeczno-ekonomicznych gospodarstw i ich członków,
- weryfikacja narzędzi badawczych obejmująca:
 - ✓ ocenę sposobu konstrukcji kwestionariuszy gospodarstwa domowego i członków gospodarstwa domowego podlegających badaniu,
 - ✓ ocenę poszczególnych, zawartych w kwestionariuszach pytań (poprawność i kompletność pytań, dopasowanie pytań do możliwości odpowiedzi na nie przez respondentów, stopień trafności i istotności odpowiedzi na nie przez respondentów),
- sprawdzenie aspektów organizacyjnych i technicznych badania (sprawdzenie jakości sieci ankieterów, czasu realizacji wywiadów, efektywności i sposobu aranżacji wywiadów, efektywności wstępnej rekrutacji do panelu telemetrycznego, jakości i kompletności kontroli realizacji badania przez sieci ankieterskie je realizujące, poprawności konstrukcji i kompletności bazy wyników uzyskanych w badaniu oraz czasu jej konstrukcji),
- ocenę poprawności konstrukcji instrukcji dotyczącej badania.
- Badanie pilotażowe było prowadzone przez trzy firmy zewnętrzne.



Realizacja badania
w terenie

6. Realizacja badania w terenie

6.1. Technika zbierania danych

Zarówno badanie pilotażowe jak i Badanie Założycielskie były przeprowadzone techniką CAPI (*Computer Assisted Personal Interviewing*), czyli dane były zbierane na podstawie wywiadu bezpośredniego (*face-to face*) w asyście laptopa/tabletu, na którym została zainstalowana aplikacja ankietera dostarczona przez KIM.

Przy wyborze techniki zbierania danych był brany pod uwagę jej wpływ na realizację zamierzonych celów badania. Od wyboru techniki badania zależy także wprost dopuszczalny sposób konstrukcji i losowania próby. Badanie ankietowe gospodarstw domowych, którego wyniki mają być uogólniane na całą populację z podaniem błędów szacunku jej parametrów powinno być oparte na próbie losowej. Taką możliwość daje jedynie badanie przeprowadzone techniką CAPI, w którym próba była losowana z operatu losowania opartego o TERYT. Jednocześnie ten operat losowania zapewnia najniższy błąd związany z doбором próby (*sampling error*) oraz z pokryciem badanej populacji (*coverage error*). Ponadto tylko wizyta w mieszkaniu gospodarstwa domowego pozwala na spis z natury wyposażenia gospodarstw domowych i jego członków w urządzenia do odbioru mediów elektronicznych oraz sprawdzenia gotowości gospodarstw domowych na nowy standard odbioru telewizji, rozszerzenia możliwości naziemnej telewizji cyfrowej oraz połączenia odbiorników telewizyjnych z internetem. Technika CAPI umożliwia uzyskanie najniższego odsetka odmów udziału w badaniu, jak również najniższego odsetka odmów odpowiedzi na pojedyncze pytania. W efekcie badanie to charakteryzuje się najniższym błędem braku odpowiedzi (*non-response error*). Wreszcie badanie prowadzone tą techniką minimalizuje błąd pomiaru (*measurement error*). Osobisty kontakt z ankieterem umożliwia również respondentowi lepsze zrozumienie pytań.

W przypadku niedostępności członków gospodarstwa domowego lub braku czasu na zrealizowanie wywiadu w trakcie wizyty w wylosowanym mieszkaniu ankieter ustalał termin powtórnego kontaktu i zbierał dane kontaktowe do tych osób celem przeprowadzenia wywiadu telefonicznego.



6.2. Sieć ankieterska

Badanie Założycielskie było realizowane przez cztery sieci ankieterskie: trzech wykonawców wybranych w drodze przetargu oraz własną sieć ankieterską KIM. Dywersyfikacja wykonawców badania umożliwiła weryfikację jakości pracy ich sieci ankieterskich oraz zminimalizowanie efektu ankieterskiego. Ponadto zaangażowanie w realizację badania bardzo dużej liczby ankieterów pozwoliło na jego realizację w terenie w założonym czasie, pomimo utrudnień związanych z pandemią.

6.3. Sposób realizacji próby

Adresy wylosowane do próby podstawowej zostały podzielone na transze przeznaczone do realizacji przez poszczególne sieci ankieterskie. Każda z transz próby adresów miała identyczną strukturę, zgodną ze strukturą ogólnopolską, zapewniając ich reprezentatywność w skali kraju. W każdej wiązce ankieter otrzymywał po 20 ponumerowanych adresów. Adresy 1-4 stanowiły próbę zasadniczą, a następane numery określały kolejność podejmowania prób kontaktów w przypadku niemożności zrealizowania wywiadu pod adresami wskazanymi przy niższych numerach (próba rezerwowa). W sytuacjach niezrealizowania wywiadów mimo użycia wszystkich 20 przekazanych adresów KIM, po przeanalizowaniu przyczyn tych sytuacji, podejmował decyzje odnośnie do potrzeby wylosowania dodatkowych adresów. W przypadku podjęcia takiej decyzji losowano w tych samych rejonach statystycznych odpowiednią liczbę dodatkowych adresów rezerwowych. Dzięki zastosowaniu adresów rezerwowych w tych samych rejonach statystycznych struktura próby zrealizowanej odpowiadała strukturze próby wylosowanej, co jest istotne ze względu na konieczność równomiernego pokrycia populacji badanej.

6.4. Wsparcie techniczne

Badanie Założycielskie jest realizowane z użyciem skryptu dostarczonego przez KIM, z wykorzystaniem Systemu Zarządzania Badaniami Instytutu. System ten wspomaga prawidłową realizację Badania Założycielskiego poprzez m.in.:

- alokację próby dla każdego wykonawcy
- kontrolowanie tożsamości ankieterów przy realizacji badania



- otrzymywanie w czasie rzeczywistym informacji dotyczących historii kontaktów, stopnia realizacji wywiadów, niezrealizowania wywiadu, wstępnych zgód respondentów na udział w winnych projektach Zamawiającego
- uzyskiwanie zdjęć budynków (wraz z ich numerami) oraz w miarę możliwości zdjęć infrastruktury mediowej
- generowania raportów:
 - ✓ raportów zbiorczych historii kontaktów ankieterów realizujących badanie pod wylosowanymi do próby adresami,
 - ✓ raportów kontroli sieci ankieterów zawierających wybrane dane zebrane przez ankieterów w trakcie wywiadów, raporty te dostępne są w systemie dla poszczególnych wykonawców badania i stanowią podstawę sporządzania przez nich cyklicznych raportów przesyłanych do KIM,
 - ✓ raportów realizacji badania, zawierających dane podsumowujące wywiady przeprowadzone przez sieć ankieterską danego wykonawcy badania, dostępne w systemie dla poszczególnych wykonawców badania.

6.5. Działania zwiększające współczynnik realizacji próby (*response rate*)

Jeden z podstawowych problemów występujących w procesie realizacji badania stanowi niepełna realizacja próby. Aby zapewnić równomierność rozkładu współczynnika realizacji według jednostek terytorialnych zastosowano uzupełnianie próby zasadniczej poprzez dodatkowe adresy losowane w uprzednio dobranych rejonach statystycznych. Jednakże mając na względzie konieczność dbania o najwyższy możliwy współczynnik realizacji próby w Badaniu Założycielskim podjęto szereg działań, a mianowicie:

- szkolenie dla ankieterów z technik zwiększających szansę uzyskania zgody na wejście na teren zamkniętego osiedla/budynku oraz z różnych sposobów aranżacji wywiadów, dla różnych grup respondentów, zachęcających ich do wzięcia udziału w badaniu oraz zasad prowadzenia wywiadu,
- wielokrotne i jednocześnie zróżnicowane próby kontaktu z respondentami, podejmowanie prób kontaktu o różnych porach dnia (przynajmniej jedna próba kontaktu w godzinach wieczornych) oraz w różnych dniach tygodnia



- możliwość przerywania badania i powtórna wizyta w umówionym terminie w celu dokończenia wywiadu lub przeprowadzenia wywiadów z innymi członkami gospodarstwa domowego,
- możliwość kontynuacji przez telefon przerwane go wywiadu indywidualnego, nie wymagającymi bezpośredniego wywiadu z respondentem,
- wręczany upominek każdemu gospodarstwu domowemu,
- bieżąca analiza przyczyn odmów pozwalająca na wprowadzanie na bieżąco działań zmierzających do zmniejszenia liczby odmów uczestnictwa w badaniu,
- przygotowanie specjalnej strony internetowej projektu, dedykowanej respondentom, pozwalającej uzyskać respondentom dodatkowe informacje o projekcie oraz zwiększające ich poczucie osobistego bezpieczeństwa; adres strony podawany jest w liście referencyjnym,
- udostępnienie respondentom specjalnej infolinii oraz mailowego helpdesku zajmujących się udzielaniem informacji i niesieniem pomocy respondentom,
- ścisły monitoring pracy ankieterów, postępów w realizacji badania i kontrola jakości wywiadów oraz szczegółowa sprawozdawczość w tych obszarach,
- ścisła i bieżąca kontrola pracy ankieterów, obejmująca zarówno przypadki braku kontaktów z respondentem jak i przyczyny odmów udziału w badaniu, budowanie motywacji ankieterów poprzez:
 - ✓ uświadomienie ankieterom udziału w ważnym projekcie naukowym,
 - ✓ podkreślanie znaczenia fazy terenowej badania w realizacji całego projektu oraz kluczowej roli ankieterów dla jej powodzenia,
- stały kontakt z ankieterami, dający możliwość zgłaszania ewentualnych problemów oraz wyjaśniania ich wątpliwości,
- optymalna liczba adresów do realizacji wywiadów (nie więcej niż 5 wiązek), pozwalająca zarówno na satysfakcjonujący ankietera zarobek jak i uwzględniający jego możliwości realizacyjne, minimalizując jednocześnie zafałszowanie odpowiedzi udzielanych przez respondenta (efekt ankieterski),
- odsuwanie od projektu ankieterów nierzetelnych i osiągających niską efektywność realizacji próby.





Kontrola
realizacji badania

7. Kontrola realizacji badania

Głównym celem kontroli realizacji badania jest zapewnienie wysokiej jakości uzyskiwanych w wyniku badania danych. Kontrola pełni zarówno funkcję walidacyjną jak i funkcję pragmatyczną. W przypadku funkcji walidacyjnej celem kontroli jest ocena badania z punktu widzenia rzetelności i wiarygodności wyników. Natomiast w przypadku funkcji pragmatycznej cel kontroli stanowi ocena realizacji badania pod kątem uczciwości i rzetelności ankietera (stwierdzenie faktu przeprowadzenia wywiadu, przeprowadzenia wywiadu z właściwymi respondentami, realizacja wywiadu zgodnie z założoną procedurą i wytycznymi zawartymi w instrukcji, itd.).

Kontrola realizacji Badania Założycielskiego prowadzona jest na trzech poziomach. Na pierwszym poziomie kontrola dokonywana jest przez skrypt wywiadu i obejmuje sprawdzenie kompletności i poprawności wypełnienia każdego kwestionariusza. Ankieter dodatkowo uzupełnia historię kontaktów pod wylosowanym adresem.

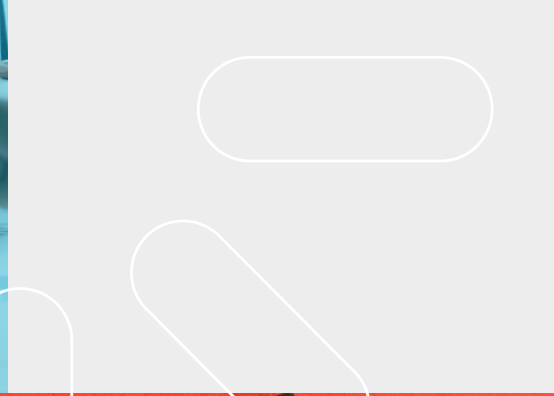
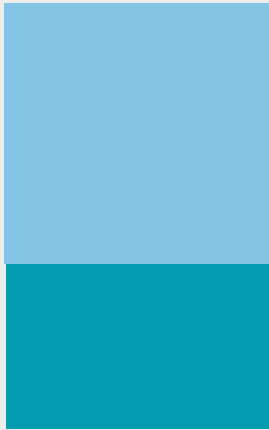
Drugi etap kontroli realizowany jest na poziomie wykonawców realizujących badanie w terenie. Wykonawcy badania kontrolują pracę podległych ankieterów obejmując kontrolą terenową minimum 10% wywiadów oraz 100% kontrolą merytoryczną. Wykonawcy cyklicznie sporządzają raport z wynikami kontroli i przesyłają go do KIM.

Trzeci poziom kontroli to kontrola prowadzona przez KIM, która ma na celu weryfikację danych zebranych podczas realizacji badania wywiadów pod kątem poprawności logicznej i rzetelności pracy ankieterów. Kontrola ta ma charakter dwustopniowy. Wewnętrzna kontrola merytoryczna jest prowadzona w sposób ciągły przez dedykowany zespół kontroli KIM w oparciu o pliki zawierające odpowiedzi respondentów na pytania zawarte w ankietach, informacje ewidencyjne (m.in. czas trwania wywiadu, datę wywiadu, adres itd.) oraz pliki z historią kontaktów pod wylosowanym adresem. Wyniki bieżącej analizy przebiegu badania pozwalają zarówno na eliminację ankieterów nierzetelnych i nieaktywnych, jak i wytypowanie wywiadów wymagających kontroli terenowej.

Kontrola terenowa jest prowadzona przez zewnętrzne, niezależne podmioty wyłonione w drodze przetargu. Obejmuje ona 10% wylosowanych do próby adresów, w tym zarówno



wywiadów efektywnych jak i nieefektywnych. Lista kontrolowanych adresów jest tworzona przez KIM. Obejmuje ona zarówno wywiady, których wyniki wzbudziły wątpliwości w trakcie przeprowadzonej wewnętrznej kontroli merytorycznej, jak i wywiady dobrane do kontroli w sposób losowy. Zewnętrzne firmy przeprowadzają kontrolę w oparciu o skrypt kontrolny, który jest opracowany na podstawie kwestionariusza kontroli dostarczonego przez KIM. Kontrola terenowa jest prowadzona techniką wywiadu osobistego lub wywiadu telefonicznego. na podstawie przeprowadzonych kontroli zewnętrznych, firmy realizujące kontrolę terenową sporządzają raporty przesyłane do KIM. Wyniki tej kontroli przekazywane są Wykonawcom oraz wykorzystywane są następnie w kontroli wewnętrznej. Kontrola zewnętrzna terenowa, realizowana przez niezależne podmioty gwarantuje dodatkową weryfikację zebranego materiału empirycznego oraz kontroli wewnętrznej.

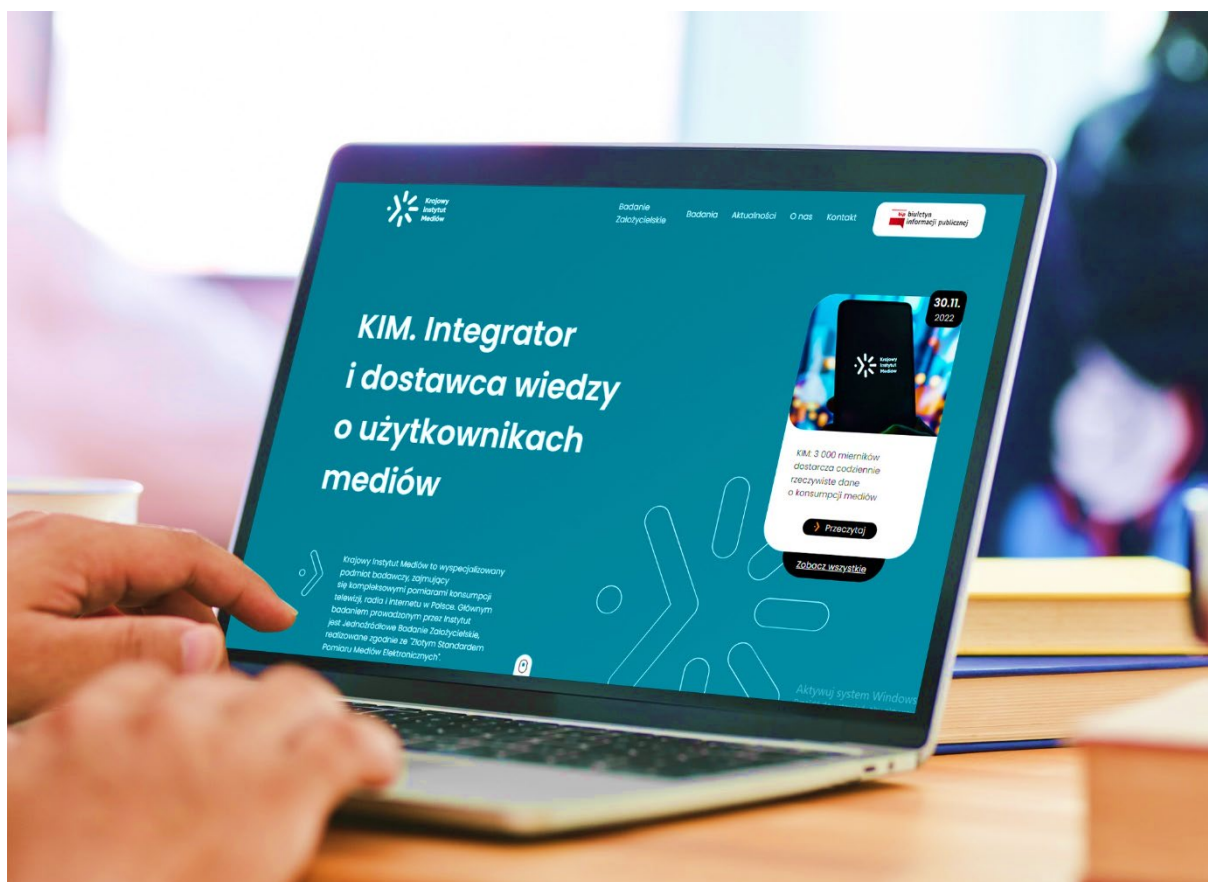


Przetwarzanie
danych

8. Przetwarzanie danych

Dane z przeprowadzonych przez ankieterów wywiadów z gospodarstwami domowymi i ich członkami są automatycznie zapisywane w Systemie Zarządzania Badaniami KIM. Dane te są następnie eksportowane na serwer KIM tworząc operacyjną bazę mikrodanych.

W bazie tej uruchamiane są procesy mające na celu sprawdzenia poprawności danych, pod kątem ich spójności oraz zgodności z próbą wylosowaną do Badania Założycielskiego. Finalnym produktem jest oczyszczony zbiór danych, który po odpersonalizowaniu (usunięciu danych pozwalających na identyfikację zbadanych gospodarstw domowych) tworzy analityczną bazę mikrodanych. w analitycznej bazie danych dokonywane są dalsze przekształcenia danych (np. ważenie danych) oraz przeprowadzane są wszelkiego rodzaju analizy. Wyniki wybranych analiz prezentowane są cyklicznie na stronie internetowej KIM.





Konstrukcja systemu wag
- uogólnianie wyników
na badaną populację

9. Konstrukcja systemu wag – uogólnianie wyników na badaną populację

Dla uogólniania uzyskanych na podstawie próby wyników Badania Założycielskiego na badaną populację w układzie ogólnopolskim oraz na podpopulacje w wybranych przekrojach typologicznych gospodarstw domowych oraz osób zastosowano zintegrowany system wag. Wagi te mają na celu przywrócenie w trakcie obliczeń pierwotnej struktury próby, zniekształconej odmowami uczestnictwa w badaniu (odmowami badanych gospodarstw i ich członków). Ważenie ma także na celu skorygowanie rozkładów cech z próby (tak gospodarstw domowych jak i osób) w oparciu o dane dostępne z niezależnych i wiarygodnych źródeł o rozkładach tych cech dla populacji. Tego typu ważenie eliminuje błędy losowe tkwiące w dobranej próbie.

Przy szacunku tych wag były uwzględniane:

- prawdopodobieństwa wyboru do próby mieszkań i gospodarstw domowych,
- poziom kompletności badania,
- struktura demograficzna populacji gospodarstw domowych i osób w populacji generalnej na podstawie danych pochodzących z zewnętrznych źródeł dostępnych w momencie tworzenia wag (dla gosp. domowych: prognoza demograficzna GUS na podstawie NSP 2011; dla osób: NSP 2021 wg stanu z czerwca 2023 uzupełniony o dane z banku danych lokalnych GUS oraz pomocniczo badania BAEL GUS).

W ramach badania zostały obliczone następujące podstawowe wagi przekrojowe:

- dla gospodarstw domowych,
- dla osób.

W kolejnym kroku zastosowano metodę zintegrowanego ważenia (tzw. procedura zintegrowanej kalibracji, Estevao i Särndal, 2006). Celem tej procedury jest zwiększenie precyzji estymatorów wykorzystując zewnętrzne źródła informacji we wzorze na dany estymator.

Ideą metody zintegrowanej kalibracji jest jednoczesny szacunek wag dla gospodarstw domowych i ich członków. Jako zmienne kalibracyjne dla badania założycielskiego zostały wykorzystane następujące zmienne:

- na poziomie gospodarstwa domowego wykorzystano informacje o liczbie gospodarstw domowych według wielkości gospodarstwa domowego (wyróżniono 5 kategorii wielkości: 1-osobowe, 2-osobowe, 3-osobowe, 4-osobowe, 5 osobowe i więcej (5 zmiennych kalibracyjnych) oraz o liczbie gospodarstw domowych według województw (16 zmiennych kalibracyjnych),
- na poziomie osób wykorzystano informacje o płci, miejscu zamieszkania (miasto/wieś), grupie wieku osób (wyróżnionych zostało 9 grup wieku: 0-3, 4-6, 7-15, 16-19, 20-29, 30-49, 50-64, 65-74 oraz 75+) oraz wykształceniu (wykształcenie wyższe, wykształcenie średnie, wykształcenie niższe niż średnie, przy czym dla osób z pierwszych trzech grup wieku, tj. poniżej 20 r. ż. nie wyróżniano poziomu wykształcenia).

Łącznie wyróżniono więc 21 komórek kalibracyjnych dla gospodarstw domowych (16+5) oraz 76 komórek dla osób ($2 \times 2 \times 4 = 16$ dla osób do 20 r.ż. oraz $2 \times 2 \times 5 \times 3 = 60$ dla osób w wieku 20 lat i więcej).

Informacje dotyczące zmiennych kalibracyjnych dla osób pochodziły z Narodowego Spisu Powszechnego Ludności i Mieszkań 2021, uzupełnione o dane z banku danych lokalnych GUS oraz pomocniczo badania BAEL GUS. Informacje o liczbie gospodarstw domowych zostały zaczerpnięte z prognoz dla 2023 roku wykonanych przez GUS na podstawie spisu 2011 roku aktualizowanych o dane rejestrowe. Oparcie liczby mieszkań o starszy spis spowodowane było odmienną definicją gospodarstwa domowego w badaniu założycielskim (definicja ekonomiczna) oraz spisie powszechnym 2021 (definicja adresowa).



Ocena precyzji estymacji

- uogólnianie wyników
na badaną populację

10. Ocena precyzji estymacji - uogólnianie wyników na badaną populację

Wyniki Badania Założycielskiego, tak jak wyniki każdego badania opartego na próbie, są obciążone błędem losowym (*sampling error*). nie wynika on z żadnych zaniedbań, czy też przypadkowych błędów badawczych. Powstaje on wyłącznie w powodu poddania badaniu losowej próby członków badanej populacji zamiast całej populacji, a wylosowana próba nie jest idealną reprezentacją populacji, z której została wylosowana.

Na podstawie uzyskanych danych indywidualnych, z wykorzystaniem załączonego do zbioru danych systemu wag, można teoretycznie oszacować parametry z populacji generalnej dla dowolnych przekrojów terytorialnych i grup typologicznych gospodarstw domowych oraz respondentów. Należy jednak mieć świadomość, że dla przekrojów, w ramach których liczebności próby są niewielkie, wielkość błędów losowych oszacowań parametrów może być znacząca a tym samym wnioskowanie na podstawie oszacowań z próby obarczone dużym błędem. Aby ocenić wiarygodność oszacowań parametrów z populacji generalnej dla danego przekroju populacji należy obliczyć dla każdego z tych oszacowań odpowiadające im błędy szacunku.

Estymatorem dla średniej wartości zmiennej (cechy) X w populacji generalnej mieszkań jest średnia ważona z próby:

$$\hat{Y} = \frac{1}{M} \sum_{w=1}^{16} \sum_{h=1}^9 \sum_{i=1}^{m_{wh}} \sum_{j=1}^{n_{whi}} x_{whij} W_{whij} \quad (1)$$

gdzie:

M – liczba mieszkań w populacji,

x_{whij} – wartość zmiennej X dla j -tej jednostki z i -tego mieszkania (j -tego gospodarstwa domowego albo j -tego jego członka zamieszkującego w i -tym mieszkaniu) w h -tej klasie miejscowości zamieszkania w -tego województwa,

w_{whij} – waga j -tej jednostki zamieszkującej i -te mieszkanie, w h -tej klasie miejscowości zamieszkania w -tego województwa,

n_{whi} – liczba jednostek w i -tym mieszkaniu w h -tej klasie miejscowości zamieszkania w -tego województwa,

m_{wh} – liczba mieszkań w h -tej klasie miejscowości zamieszkania w -tego województwa.



W przypadku gdy zmienna X jest zmienną zero-jedynkową średnia wartość tej zmiennej określa frakcję gospodarstw domowych albo osób posiadających dany wariant zmiennej w populacji generalnej.

Standardowy błąd szacunku estymatora (1), czyli ocenę jego odchylenia standardowego, dla populacji generalnej uzyskujemy według wzoru:

$$ASE(\hat{Y}) = \sqrt{V(\hat{Y})} = \sqrt{\frac{1}{M^2} \sum_{w=1}^{16} \sum_{h=1}^9 \frac{m_{wh}}{m_{wh}-1} \sum_{i=1}^{m_{wh}} (x_{whi} - \bar{x}_{wh})^2}, \quad (2)$$

gdzie:

x_{whi} – ważona suma wartości zmiennej X w i -tym gospodarstwie domowym w h -tej warstwie w -tego województwa, przy czym:

$$x_{whi} = \sum_{j=1}^{n_{whi}} x_{whij} \cdot w_{whij}, \quad (3)$$

\bar{x}_{wh} – średnia wartość z sumy wartości zmiennej X w h -tej warstwie, w -tego województwa, przy czym:

$$\bar{x}_{wh} = \frac{1}{m_{wh}} \sum_{i=1}^{m_{wh}} x_{whi}. \quad (4)$$

Błędy szacunku parametrów dla wyróżnionych przekrojów typologicznych gospodarstw domowych i osób są szacowane na podstawie analogicznej procedury.



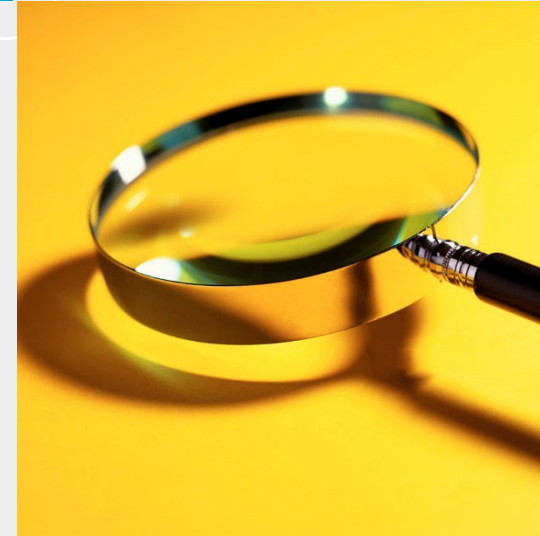
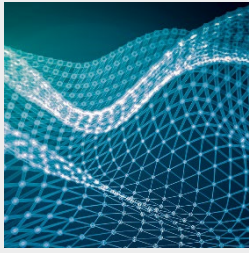
Udostępnianie wyników badania



11. Udostępnianie wyników badania

Wyniki z Badania Założycielskiego prezentowane są w postaci raportów i analiz ogólnodostępnych na stronie internetowej KIM. Raporty i analizy publikowane w cyklu półrocznym i rocznym, obejmując wyniki kolejnych etapów tego badania.

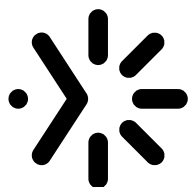




Szacunek uniwersów

12. Szacunek uniwersów

Jednym z celów Badania Założycielskiego jest szacunek uniwersów, tj. liczebności gospodarstw domowych i osób według ich cech różniących konsumpcję usług medialnych, które są niedostępne w zbiorach danych statystyki publicznej. Szacunki uniwersów są wykorzystywane do ważenia gospodarstw domowych i ich członków według ich cech demograficznych i społeczno-ekonomicznych stosowanych w analizach wyników Badania Założycielskiego oraz są wykorzystywane do celów sprawozdawczych i wykorzystywane do ustalania prawidłowej struktury próby w innych badaniach. Dotychczasowe doświadczenia wskazują, że aktualizacja uniwersów powinna być prowadzona raz w roku.



Krajowy
Instytut
Mediów



KRAJOWY INSTYTUT MEDIÓW

ul. Wiktorska 63,
02-587 Warszawa

KONTAKT

E-mail:

biuro@kim.gov.pl

Telefon:

22 118 98 44
(w godz. 9.00-17.00)

Wszelkie prawa zastrzeżone dla Krajowego Instytutu Mediów. Jakiegokolwiek dalsze przetwarzanie, rozpowszechnianie, modyfikowanie, korzystanie z treści i informacji przedstawionych przez Krajowy Instytut mediów oraz Krajową Radę Radiofonii i Telewizji – w całości lub w postaci opracowanej części, z bazy danych oraz utworów powstałych w ramach badania – wymaga uzyskania stosownej licencji od Krajowego Instytutu Mediów.



KRAJOWY INSTYTUT MEDIÓW