



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI



IBE  *entuzjaści  
edukacji*

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



*Uwarunkowania  
Decyzji  
Edukacyjnych*

# UWARUNKOWANIA DECYZJI EDUKACYJNYCH

## ZASADY KONSTRUKCJI PRÓBY PANELOWEJ I ZASADY WNIOSKOWANIA STATYSTYCZNEGO

Materiał opracowany przez zespół ekspertów Szkoły Głównej Handlowej w Warszawie na zlecenie Instytutu Badań Edukacyjnych w ramach realizowanego projektu systemowego pod nazwą „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego” współfinansowanego ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki

**Warszawa, 2014**



# Spis Treści

<b>1. Wprowadzenie .....</b>	<b>4</b>
<b>2. Jednostka badania .....</b>	<b>4</b>
<b>3. Założenia panelu .....</b>	<b>5</b>
<b>4. Zasady definiowania próby panelowej.....</b>	<b>7</b>
4.1. Zasady tworzenia początkowej próby (podpróbek) gospodarstw domowych oraz identyfikacji osób tworzących panelową próbę osób .....	7
4.2. Podstawowe reguły określania statusu osób podlegających badaniu panelowemu .....	7
4.3. Zasady identyfikacji gospodarstw domowych podlegających badaniu panelowemu .....	8
4.4. Zasady identyfikacji osób podlegających badaniu w kolejnych rundach panelu .....	9
4.5. Podążanie za osobami należącymi do panelowej próby – reguły tzw. „śledzenia” .....	10
4.5.1. Przemieszczanie się osób .....	10
4.5.2. Przedział wieku .....	10
4.5.3. Braki odpowiedzi .....	10
4.6. Szczegółowe zasady procedury śledzenia .....	10
4.7. Longitudinalna identyfikacja elementów próby panelowej .....	12
4.7.1. Numery identyfikacyjne gospodarstw domowych .....	13
4.7.2. Numery indentyfikacyjne osób .....	14
<b>5. Próba oraz uogólnianie wyników badania na populację .....</b>	<b>14</b>
5.1. Założenia.....	14
5.2. Schemat losowania próby .....	15
5.2.1. Schemat losowania próby na poziomie ogólnopolskim .....	16
5.2.2. Schemat losowania próby na szczeblu lokalnym .....	17
<b>6. Ustalenie minimalnej wielkości próby .....</b>	<b>18</b>
<b>7. Próby rezerwowe .....</b>	<b>19</b>
<b>8. Wagi .....</b>	<b>20</b>
8.1. Wagi przekrojowe .....	20
8.2. Wagi longitudinalne .....	25

<b>9. Metoda oceny precyzji estymacji .....</b>	<b>25</b>
<b>Literatura.....</b>	<b>26</b>

# 1. Wprowadzenie

Zrealizowanie przyjętych w projekcie zamierzeń badawczych wymagało opracowania i zastosowania takiej metodyki badawczej, która umożliwiałaby zbieranie informacji niezbędnych do prowadzenia zarówno analiz przekrojowych, jak i analiz zmian w czasie na poziomie ogólnokrajowym i regionalnym oraz na poziomie lokalnym. W związku z tym zdecydowano, by badanie **Uwarunkowań decyzji edukacyjnych** zostało zrealizowane metodą panelu.

Metodyka badania zapewnia integrację badania panelowego gospodarstw domowych i ich dorosłych członków na szczeblu ogólnokrajowym i regionalnym oraz lokalnym co wymagało:

- zdefiniowania próby panelowej;
- opracowania zasad prowadzenia panelowej próby gospodarstw domowych i osób;
- opracowania metod szacunku parametrów z populacji generalnej (cech badanych jednostek i wskaźników), w tym systemu wag związanych z uogólnianiem wyników badania na populację, określenia charakteru próby i planu jej losowania, schematu losowania próby i minimalnej wielkości próby oraz sformułowania założeń operatu losowania;
- opracowania zasad oceny precyzji wyników.

Badanie krajowe ma charakter ogólnopolski, a liczebność i sposób konstrukcji próby (plan losowania i schemat losowania) oraz zastosowany system wag gwarantują możliwość szacowania wartości wiodących cech z dostateczną precyzją, na poziomie ogólnokrajowym (NUTS0) oraz województw (NUTS2, poziom regionalny). Natomiast badanie na poziomie lokalnym umożliwia szacowanie wartości cech wiodących z dostateczną precyzją na poziomie powiatów (NUTS3). Wyniki pomiaru umożliwiają ponadto analizę danych zarówno w podejściu przekrojowym, jak i longitudinalnym.

## 2. Jednostka badania

W projekcie przyjęto, że jednostką badania jest gospodarstwo domowe oraz wszyscy jego członkowie, którzy do dnia 31 grudnia 2012 r. nie ukończyli jeszcze 65 lat. Za gospodarstwo domowe przyjęto, zgodnie z obowiązującą w polskiej statystyce definicją (GUS, Metodologia..., 2011), zespół osób spokrewnionych ze sobą lub niespokrewnionych, mieszkających razem i wspólnie utrzymujących się (gospodarstwo domowe wieloosobowe) lub osobę utrzymującą się samodzielnie, bez względu na to, czy mieszka sama (gospodarstwo domowe jednoosobowe), czy też z innymi osobami. Członkowie rodziny mieszkający wspólnie, ale utrzymujący się oddzielnie, tworzą odrębne gospodarstwa domowe. Wielkość gospodarstwa domowego jest określana liczbą osób wchodzących w jego skład.

Do składu gospodarstwa domowego zaliczono:

- osoby mieszkające razem w gospodarstwie domowym i wspólnie utrzymujące się, przebywające w gospodarstwie domowym przez okres co najmniej sześciu miesięcy (należy wziąć pod uwagę rzeczywisty lub zamierzony okres przebywania w gospodarstwie domowym);
- osoby nieobecne w gospodarstwie domowym z racji pracy zawodowej, jeśli dochody z pracy tych osób przekazywane są na utrzymanie gospodarstwa;

- osoby nieobecne w gospodarstwie domowym, w wieku do 15 lat włącznie, pobierające naukę poza miejscem zamieszkania, mieszkające w internatach lub w domach prywatnych;
- osoby nieobecne w gospodarstwie domowym w okresie badania, przebywające w zakładach opiekuńczo-wychowawczych, w domach opieki lub szpitalach, jeżeli ich rzeczywisty lub zamierzony okres pobytu poza gospodarstwem domowym jest krótszy niż 6 miesięcy.

Do składu gospodarstwa domowego nie zaliczono:

- osób nieobecnych w gospodarstwie domowym, w wieku powyżej 15 lat, pobierających naukę poza miejscem zamieszkania, mieszkających w internatach, domach akademickich lub w domach prywatnych; osoby te tworzą odrębne gospodarstwa domowe w miejscach swojego pobytu;
- mężczyzn odbywających zasadniczą służbę wojskową (poborowi odbywający służbę zastępczą w zakładach pracy i mieszkający w domu wchodzi w skład gospodarstwa domowego);
- osób przebywających w zakładzie karnym;
- osób nieobecnych w gospodarstwie domowym w okresie badania, przebywających w zakładach opiekuńczo-wychowawczych, w domach opieki lub w szpitalach, jeżeli ich rzeczywisty lub zamierzony okres pobytu poza gospodarstwem domowym jest dłuższy niż 6 miesięcy;
- osób przebywających w gospodarstwie domowym w okresie badania (gości), obecnych w gospodarstwie domowym lub zamierzających w nim przebywać przez okres krótszy niż sześć miesięcy;
- lokatorów, w tym uczniów i studentów na stacji (z wyjątkiem sytuacji, gdy są oni traktowani przez badane gospodarstwo domowe jako członkowie gospodarstwa);
- osób wynajmujących pokój lub łóżko, na okres pracy w danej miejscowości (dotyczy to np. takich prac jak melioracje, pomiar gruntów, wyrąb lasów, budowa);
- osób mieszkających wspólnie z gospodarstwem domowym, zatrudnionych przez to gospodarstwo, np. pomocy domowych, robotników rolnych, uczniów i praktykantów w zawodzie.

W ramach projektu, badaniu podlegają prywatne gospodarstwa domowe, pomijane są zaś 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).

### 3. Założenia panelu

Podstawową cechą panelu proponowanego do zastosowania w badaniu (zarówno na poziomie ogólnokrajowym jak i na poziomie lokalnym) jest zintegrowany schemat losowania zapewniający, że wartości statystyk przekrojowych (dla danego roku) i statystyk longitudinalnych (dla wielu lat) otrzymywane są z tego samego, w zasadniczej części, zbioru obserwacji z próby. Przyjęte rozwiązanie umożliwia uniknięcie prowadzenia dwóch niezależnych badań, o charakterze przekrojowym i o charakterze longitudinalnym (zmian w czasie).

W praktyce możliwy jest różny sposób prowadzenia panelu. W projekcie zrealizowano wariant panelu klasycznego, zakładający dołosowywanie podpróbki do drugiej rundy w przypadku utraty członków z panelu początkowego. Oznacza to, że w momencie rozpoczęcia badania wylosowana została próba, która jest reprezentatywna ze względu na jej strukturę z całą populacją. W kolejnej rundzie badaniem

objęte są wszystkie gospodarstwa domowe (i ich członkowie) ułożone w wylosowywanych do próby mieszkaniach, a w przypadku utraty członków z próby początkowej dołosowywana jest próba dopełniająca wielkość próby wyjściowej do założonych rozmiarów. Schemat dołosowywania podpróbki przedstawia tabela 1.

**Tabela 1.** Schemat dołosowywania podpróbek do badania w panelu klasycznym w latach 2012-2014

Rok badania	Numer podróbki	
2013	1	
2014	1	2

Tylko gospodarstwa domowe i ich członkowie, którzy uczestniczą w badaniu we wszystkich rundach, tworzą tzw. początkową próbę panelową (podpróbka o numerze 1 ze schematu przedstawionego w tablicy 1 tworzy panel dwurundowy). Próba panelowa jest zatem częścią próby objętej badaniem w kolejnych rundach. Próbę panelową osób tworzą osoby w wieku od 15 do 65 lat, należące do gospodarstw domowych tworzących próbę początkową. Pozostałe osoby, nie należące do panelowej próby osób, nazywane są współmieszkańcami. Gospodarstwo domowe należy do panelowej próby gospodarstw domowych, jeżeli należy do niego przynajmniej jedna osoba należąca do panelowej próby osób.

Łączne dane dla danego roku są wykorzystywane do analiz przekrojowych dla tego roku. Operowanie w każdym roku w analizach przekrojowych pełną próbą (wszystkimi dołosowywanymi podpróbkami łącznie) pozwala na szacunek statystyk nie tylko dla całej badanej populacji, ale również dla różnych przekrojów gospodarstw domowych i osób. Jednocześnie od drugiej rundy badania są do dyspozycji dane z panelu gospodarstw domowych, które mogą być wykorzystywane do analiz zmian w czasie.

Podstawowym celem longitudinalnego elementu badania jest obserwacja zmian na poziomie pojedynczych gospodarstw domowych (na przykład zmiany w poziomie dochodów) oraz ich członków (na przykład przepływ osoby ze szkoły do pracy). Jednym z najważniejszych celów, jaki ma pełnić ten element badania, jest możliwość oceny trwałości pewnych zjawisk i procesów, co w konsekwencji wymaga obserwacji gospodarstw domowych i ich członków przez co najmniej dwie rundy badania.

Założono, że w każdej rundzie badania longitudinalny składnik badania będzie reprezentować całą badaną populację, obejmując gospodarstwa domowe oraz osoby będące ich członkami.

W kolejnych rundach badania będą stosowane reguły tzw. „śledzenia”. Reguły te zagwarantują prowadzenie obserwacji w drugiej rundzie badania wszystkich jednostek, które zaliczono do początkowej próby (gospodarstw domowych i ich członków), a to umożliwi odnotowanie wszystkich zmian, jakie zachodzą w czasie w badanej populacji.

## 4. Zasady definiowania próby panelowej

Obserwowana panelowa próba gospodarstw domowych stanowi pewien dynamicznie zmieniający się wycinek populacji polskich gospodarstw domowych. Przyjęto założenie o nieuzupełnianiu panelowej próby gospodarstw domowych w kolejnych rundach badania, gdy gospodarstwa domowe z panelowej próby ulegają naturalnemu wymarciu lub odmawiają dalszego uczestnictwa w badaniu. Pierwszą z tych sytuacji traktuje się jako naturalne wymieranie części populacji gospodarstw domowych. Natomiast w drugim przypadku, aby ubytki z gospodarstw domowych nie miały wpływu na ocenę dynamiki zmian zjawisk i procesów, zamiast zastępowania gospodarstw domowych, które wypadły z próby panelowej, zaproponowano odpowiedni system ważenia wyników. Jednocześnie do wyjściowej panelowej próby gospodarstw domowych są dołączane w drugiej rundzie badania nowe gospodarstwa domowe, do których przenieśli się członkowie gospodarstw domowych należących do wyjściowej panelowej próby gospodarstw domowych, czyli powstające na skutek podziału wyjściowej panelowej próby gospodarstw.

Dynamiczny sposób traktowania próby panelowej wymaga nie tylko zdefiniowania wyjściowych prób panelowych zarówno gospodarstw domowych (panelowej próby gospodarstw domowych), jak i ich członków (panelowej próby osób), ale także ustalenia reguł traktowania tych jednostek badania w kolejnych jego rundach.

### 4.1. Zasady tworzenia początkowej próby (podpróbek) gospodarstw domowych oraz identyfikacji osób tworzących panelową próbę osób

Informacje zbierane w ramach badania odnoszą się do następujących zbiorowości jednostek:

- prywatne gospodarstwa domowe, w celu zebrania informacji o wartościach zmiennych na poziomie gospodarstwa;
- wszyscy członkowie gospodarstw, w celu zebrania informacji o charakterze demograficznym i innych podstawowych informacji o członkach badanych gospodarstw domowych, włącznie z wielkością gospodarstwa domowego i jego wieloraką strukturą;
- członków gospodarstw domowych w wieku do 65 lat, w celu zebrania informacji o charakterze szczegółowym odnoszących się do ich aktywności edukacyjnej, rodzinnej, zawodowej i migracyjnej;
- osób, które były członkami badanych gospodarstw domowych w poprzedniej rundzie panelu (lecz w danej rundzie badania już nimi nie są), w celu zebrania podstawowych informacji oraz odnotowania czasu ich przebywania w gospodarstwie domowym w okresie międzyrundowym. Informacje te są zbierane tylko dla elementu longitudinalnego próby.

### 4.2. Podstawowe reguły określania statusu osób podlegających badaniu panelowemu

Wywiadom indywidualnym, prowadzonym bezpośrednio z respondentem, podlegają wszystkie osoby należące do panelowej próby osób w wieku 15-65 lat. Ponadto, od drugiej rundy wywiadowi



indywidualnemu będą podlegały dzieci, urodzone przez kobiety należące do panelowej próby osób, gdy ukończą 15 lat. W ten sposób zbiorowość osób podlegających badaniu panelowemu zostaje uaktualniana w kolejnej rundzie panelu, zgodnie z zachodzącymi zmianami demograficznymi.

Do osób, które w kolejnej rundzie podlegają wywiadam indywidualnym lecz nie są członkami panelowej próby osób, należą wszystkie osoby w wieku 15-65 lat, które w drugiej rundzie badania tworzą gospodarstwa domowe z przynajmniej jedną osobą należącą do panelowej próby osób, czyli weszły w skład gospodarstw domowych podlegającym badaniu. Są to tym samym osoby, które nie były członkami gospodarstw domowych tworzących panelową próbę gospodarstw i objętych badaniem w pierwszej rundzie (początkowy panel gospodarstw domowych), lecz stały się ich członkami w drugiej rundzie panelu. W sytuacji, gdy osoby te przeniosły się do gospodarstw domowych, w których żaden z członków nie należy do panelowej próby osób, to wypadają z dalszego badania.

Przedstawione reguły określania zbiorowości osób podlegających badaniu panelowemu w drugiej rundzie wymagają definiowania ich aktualnego „statusu” wynikającego z ich statusu w poprzedniej rundzie badania.

### 4.3. Zasady identyfikacji gospodarstw domowych podlegających badaniu panelowemu

Zgodnie z opisanymi wyżej zasadami, wszystkie gospodarstwa, które uczestniczyły w pierwszej rundzie badania, tworzą panelową próbę gospodarstw domowych. Gospodarstwa domowe nie podlegające badaniu zostają usunięte z panelowej próby gospodarstw. Zasady identyfikacji gospodarstw domowych, podlegających badaniu w pierwszej rundzie w oparciu o ich status w badaniu w obu rundach, przedstawiono w tabeli 2.

**Tabela 2.** Zasady identyfikacji gospodarstw domowych należących do panelowej próby gospodarstw w rundzie R-1

Status gospodarstwa domowego w drugiej rundzie	Status gospodarstwa domowego w pierwszej rundzie		
	Przeprowadzony wywiad	Nieprzeprowadzony wywiad z powodu: niemożności brania udziału w badaniu (np. podeszły wiek, choroba) braku kontaktu, wstępnej odmowy	Nieprzeprowadzony wywiad z powodu: definitywnej odmowy, braku możliwości lokalizacji
Przeprowadzony wywiad	GDPB	GDPB	GDNB
Nieprzeprowadzony wywiad	GDPB	GDNB	GDNB

Nowe gospodarstwo, które nie podlegało wywiadowi w pierwszej rundzie <sup>1</sup>	GDPB	GDPB	GDNB
---	------	------	------

GDPB – gospodarstwa domowe podelegające badaniu, GDNB – gospodarstwa domowe niepodlegające badaniu (usunięte z panelowej próby gospodarstw domowych)

Z zestawienia przedstawionego w tabeli 2 wynika, że z panelowej próby gospodarstw zostają usunięte gospodarstwa, z którymi nie przeprowadzono wywiadu w dwóch kolejnych rundach panelu. Ponadto z panelowej próby gospodarstw wypadają te gospodarstwa domowe, w których na skutek zmian strukturalnych nie pozostała żadna osoba należąca do wyjściowej panelowej próby osób. Natomiast gospodarstwa domowe, których wszyscy członkowie przeszli do gospodarstw domowych zbiorowych (dotyczy to przede wszystkim jednoosobowych gospodarstw domowych), nie zostają co prawda poddawane badaniu ankietowemu w danej rundzie panelu, lecz pozostają w panelowej próbie gospodarstw. Są one poddawane tzw. procedurze „śledzenia”, umożliwiającej włączenie ich do badania ankietowego, gdy tylko staną się ponownie prywatnymi gospodarstwami domowymi. W podobny sposób traktowane są gospodarstwa domowe przenoszące się czasowo za granice kraju.

Panelowa próba gospodarstw domowych jest powiększana o gospodarstwa nowo tworzone przez osoby należące do panelowej próby osób oraz gospodarstwa, do których przenoszą się osoby należące do panelowej próby osób.

#### 4.4. Zasady identyfikacji osób podlegających badaniu w kolejnych rundach panelu

Wywiadowi indywidualnemu w danej rundzie badania podlegają wszystkie osoby w wieku 15-65 lat, należące do panelowej próby gospodarstw domowych w tej rundzie, bez względu na to czy należą do panelowej próby osób czy też do niej nie należą. Są to przede wszystkim osoby podlegające badaniu w poprzedniej rundzie, włączając osoby, z którymi nie przeprowadzono wywiadu w poprzedniej rundzie z różnych przyczyn. Ponadto, wywiadowi indywidualnemu podlegają wszystkie osoby w wieku 15-65 lat, które zostały członkami gospodarstw należących do panelowej próby gospodarstw po poprzedniej rundzie badania.

Z panelowej próby osób wypadają osoby, z którymi nie przeprowadzono wywiadu w dwóch kolejnych rundach panelu (należały do gospodarstw domowych, z którymi nie przeprowadzono wywiadu w tych rundach, czyli usuniętymi z panelowej próby gospodarstw domowych, lub też odmówiły ponownie uczestnictwa w badaniu). Wywiadowi nie podlegają także osoby należące do panelowej próby osób, które przenieśli się do gospodarstw domowych zbiorowych lub czasowo wyjechały za granicę. Nie zostają one jednakże usunięte z panelowej próby osób, lecz są „śledzone” tak, aby można było je ponownie poddać badaniu w kolejnych rundach panelu. O osobach tych zbierane są wybrane informacje (najczęściej od innych członków ich gospodarstw) dotyczące m. in. przyczyn ich czasowej nieobecności. Wreszcie wywiadowi w danej rundzie badania nie podlegają osoby nienależące do panelowej próby osób, które co prawda zostały poddane wywiadowi, gdyż należały wtedy do gospodarstw należących do panelowej próby gospodarstw, lecz następnie przenieśli się do

---

<sup>1</sup> Nowo powstałe gospodarstwo, włączone do panelowej próby gospodarstw, utworzone przez osobę należącą do panelowej próby osób lub też włączone do próby panelowej na skutek przejścia do niego przynajmniej jednej osoby należącej do panelowej próby osób.

gospodarstw, w których nie ma żadnej osoby należącej do panelowej próby osób. Osoby te wypadają z dalszego badania.

## **4.5. Podążanie za osobami należącymi do panelowej próby – reguły tzw. „śledzenia”**

Dla analizy zmian w czasie, na poziomie indywidualnym, konieczne jest podążanie w kolejnych rundach za osobami należącymi do panelowej próby osób, pomimo że mogą one zmieniać miejsce zamieszkania w czasie trwania panelu. Byłoby idealnie, gdyby wszystkie osoby należące do panelowej próby osób podlegały badaniu, bez względu na to, gdzie będą zamieszkiwały. Próba longitudinalna musi pozostać reprezentatywna dla wszystkich grup wiekowych badanej populacji. Oznacza to, że w zasadzie osoby będące członkami gospodarstw domowych należących do panelowej próby gospodarstw domowych w każdym wieku powinny podlegać śledzeniu. Jednakże w badaniu, ze względów praktycznych, zastosowano pewne ograniczenia w tym zakresie.

### **4.5.1. Przemieszczanie się osób**

W projekcie przyjęto, że śledzeniu nie podlegają te osoby, które przeniosły się do gospodarstw domowych zamieszkujących w obiektach zbiorowego zakwaterowania lub też opuściły teren kraju. Jedyny wyjątek stanowią osoby, które czasowo (na okres krótszy niż 6 miesięcy) przeniosły się do gospodarstw domowych zamieszkujących obiekty zbiorowego zakwaterowania na terenie kraju. Są one nadal uważane za członków prywatnych gospodarstw domowych i podlegają procedurze śledzenia.

### **4.5.2. Przedział wieku**

Ze względów kosztowych i praktycznych śledzenie osób zostało ograniczone do osób należących do panelowej próby osób, czyli osób w wieku 15-65 lat w momencie losowania początkowej panelowej próby gospodarstw domowych oraz podpróbki gospodarstw wchodzących do badania w następnej rundzie.

### **4.5.3. Braki odpowiedzi**

Gospodarstwa domowe, których nie udało się zlokalizować przez dwa kolejne lata badania lub też brak było z nim kontaktu w pierwszej rundzie panelu (np. z powodu braku możliwości zlokalizowania jego adresu, z powodu czasowej nieobecności wszystkich członków gospodarstwa, czy też niemożliwości przeprowadzenia wywiadu z powodu niesprawności umysłowej lub choroby) wypadają z panelu wraz ze wszystkimi jego członkami. Z próby panelowej gospodarstw domowych wypadają także gospodarstwa domowe, które co prawda wzięły udział w badaniu, lecz nie przeprowadzono w nich żadnego wywiadu indywidualnego. Jednocześnie ich członkowie, należący do panelowej próby osób, automatycznie przestają podlegać procedurze śledzenia.

## **4.6. Szczegółowe zasady procedury śledzenia**

Na podstawie przeprowadzonych powyżej ustaleń przyjęto następujące zasady śledzenia osób w badaniu:

- Dzieci w wieku poniżej 15 lat nie będą podlegały śledzeniu, jeżeli przeniosą się do gospodarstwa domowego, w którym nie będzie osób należących do panelowej próby osób. Innymi słowy, nie będą one traktowane jako osoby należące do panelowej próby osób.

- Szczegółowy wywiad będzie przeprowadzany z osobami w wieku 15-65 lat.
- Śledzeniu podlegają osoby w wieku 15-65 lat, będące członkami gospodarstw należących do panelowej próby gospodarstw, które przeniosą się do innych prywatnych gospodarstw domowych na terenie kraju.
- Śledzeniu podlegają osoby w wieku 15-65 z gospodarstw domowych należących do panelowej próby gospodarstw, przebywające czasowo w gospodarstwach domowych zamieszkujących w obiektach zbiorowego zakwaterowania. Podlegają one wywiadowi w sposób pośredni (np. na podstawie wywiadu z innymi osobami z ich gospodarstwa domowego).
- Gdy osoby w wieku 15-65 lat, należące do gospodarstw domowych z panelowej próby gospodarstw, przeniosą się do gospodarstwa zamieszkującego w obiektach zbiorowego zakwaterowania lub za granicę na stałe (na okres 6 miesięcy lub dłuższy) lub też na okres krótszy, lecz nie mogą być traktowane jako członkowie jakiegokolwiek gospodarstwa domowego, wypadają z badania (w tej sytuacji następujące informacje powinny być zebrane od członka gospodarstwa, do którego należała taka osoba w poprzedniej rundzie badania: gdzie dana osoba przeniosła się, data przeniesienia, liczba miesięcy spędzonych przez tę osobę w gospodarstwie, w okresie będącym punktem odniesienia szacowania dochodów oraz podstawowy typ aktywności zawodowej tej osoby w tym okresie).
- W stosunku do osób należących do panelowej próby osób, które zmarły, powinny być zbierane wyłącznie informacje dotyczące daty ich śmierci, liczby miesięcy spędzonych w gospodarstwie domowym w okresie, dla którego szacowane są dochody oraz podstawowego typu aktywności zawodowej w tym okresie.
- Należy spróbować nawiązać kontakt w drugiej rundzie badania z osobami w wieku 15-65 lat, należącymi do gospodarstw podlegających badaniu, z którymi z różnych przyczyn nie nawiązano kontaktu w pierwszej rundzie badania. Jeżeli nie nawiązano z tymi gospodarstwami domowymi kontaktu w pierwszym roku panelu lub też w drugiej rundzie badania, mogą zostać usunięte z badania, a wraz z nimi wszyscy ich członkowie.
- Współmieszkańcy biorą udział w badaniu tak długo jak długo są członkami gospodarstw domowych, do których należą także osoby tworzące panelową próbę osób. Będą zbierane na ich temat dane indywidualne według ogólnie przyjętych procedur, jeżeli są w wieku 15-65 lat. Jednakże współmieszkańcy nie są dalej śledzeni, jeżeli przeniosą się do gospodarstw domowych, w których nie ma osób w wieku 15 lat i więcej należących do panelowej próby osób.
- O byłych współmieszkańcach (byłych członkach gospodarstw domowych wspólnych z osobami należącymi do panelowej próby osób), którzy byli członkami badanych gospodarstw domowych przez okres przynajmniej 3 miesięcy w trakcie okresu odniesienia dla szacunku dochodów, należy zbierać następujące informacje: liczba miesięcy spędzonych w gospodarstwie domowym w okresie stanowiącym punkt odniesienia dla szacunku dochodów oraz ich podstawowy typ aktywności zawodowej w tym okresie.

Zasady postępowania w stosunku do osób należących do panelowej próby osób, gospodarstw domowych należących do panelowej próby gospodarstw oraz współmieszkańców przedstawiono w tabeli 3.

**Tabela 3.** Zasady postępowania w stosunku do osób należących do panelowej próby osób, gospodarstw domowych należących do panelowej próby gospodarstw oraz współmieszkańców.

<b>Typy jednostek</b>	<b>Sposób postępowania</b>
<b>Osoby należące do panelowej próby osób:</b> przenoszące się do prywatnego gospodarstwa domowego w kraju	śledzone w nowym miejscu zamieszkania, powinny zostać objęte badaniem
czasowo przebywające poza gospodarstwem domowym lecz jeszcze traktowane jako członkowie gospodarstwa domowego	należą dalej do swojego gospodarstwa domowego i powinny zostać objęte badaniem
nie będące dalej członkami prywatnego gospodarstwa domowego (przeprowadziły się do gospodarstwa zbiorowego lub opuściły kraj)	nie podlegają badaniu
<b>Gospodarstwa domowe należące do panelowej próby gospodarstw:</b> nie zidentyfikowane w którejś z rund ze względu na niemożliwość lokalizacji ich adresu, gdy pod danym adresem znajduje się mieszkanie niezamieszkałe, adres zamieszkania został zgubiony (brak jest informacji co stało się z danym gospodarstwem)	usunięte z badania
z którymi nie udało się nawiązać kontaktu w pierwszej rundzie panelu lub w dwóch kolejnych rundach z powodu tymczasowej nieobecności całego gospodarstwa lub niemożności przeprowadzenia wywiadu z różnych przyczyn	usunięte z badania
w całości przeprowadziło się pod inny prywatny adres w kraju	objęte badaniem w nowym miejscu zamieszkania
podzieliły się na dwa lub więcej gospodarstw domowych	objęte badaniem wszystkie gospodarstwa domowe, w skład których wchodzi przynajmniej jedna "osoba z próby panelowej" (z wyjątkiem tych gospodarstw, które przeprowadziły się za granicę lub przenieśli się do gospodarstwa zbiorowego/instytucji w kraju)
powstały z połączenia dwóch lub więcej gospodarstw domowych	objęte badaniem
<b>Współmieszkańcy:</b> należący do gospodarstwa domowego, w którym nie ma żadnej osoby należącej do panelowej próby osób	usunięci z badania

#### 4.7. Longitudinalna identyfikacja elementów próby panelowej

W badaniu panelowym jest niezwykle ważne, aby każda z osób objętych obserwacją posiadała swój unikalny numer identyfikacyjny pozwalający na powiązanie informacji o tej osobie oraz wyników wywiadów indywidualnych z tą osobą we wszystkich rundach panelu. W prowadzonych badaniach panelowych dodatkowo należy zapewnić możliwość identyfikacji w kolejnych rundach panelu

poszczególnych gospodarstw domowych oraz powiązania informacji o nich oraz wyników ich ankietowania z tych rund. Ponadto, należy uwzględnić sytuacje podziału gospodarstw domowych objętych badaniem panelowym w drugiej rundzie oraz przyjętą zasadę włączania do panelu gospodarstw, do których przeniosły się osoby należące do panelowej próby osób. Wszystko to zapewniają zasady nadawania numerów identyfikacyjnych jednostkom badania.

#### 4.7.1. Numery identyfikacyjne gospodarstw domowych

Każde gospodarstwo objęte badaniem otrzymuje swój unikalny numer identyfikacyjny, tzw. GOSPID (Identyfikator Gospodarstwa Domowego) składający się z czterech części:

N-RRRR-MMMk-PP,

gdzie:

*N* – numer rundy badania, w której dołączono mieszkanie do próby panelowej,

*RRRR* – to numer rejonu statystycznego (numer kolejny, a nie numer ze spisu GUS)

*MMM* – numer mieszkania (adresu wylosowanego w ramach rejonu spisowego)

*k* – tzw. numer gospodarstwa domowego w ramach wylosowanego mieszkania

*PP* – numer podziału gospodarstwa.

Jak już wspomniano, za gospodarstwo domowe powstałe z podziału uznaje się takie nowe gospodarstwo domowe, w skład którego wchodzi przynajmniej jedna osoba z panelowej próby osób. W tych podpróbach gospodarstw domowych, które będą badane po raz pierwszy, numer ten będzie wynosił 00. Gdy pomiędzy rundą pierwszą i drugą gospodarstwo domowe uległo podziałowi (tzn. gdy członkowie tego gospodarstwa z rundy pierwszej w rundzie drugiej znaleźli się przynajmniej w dwóch gospodarstwach, jedno z tych gospodarstw otrzymuje numer właściwy numerowi identyfikacyjnemu gospodarstwa bazowego, tj. *N - RRRR - MMMk - 00*. Kolejne gospodarstwa domowe, do których przeniosły się osoby z gospodarstwa bazowego po pierwszej rundzie badania, otrzymują ten sam numer gospodarstwa co gospodarstwo bazowe (*N - RRRR - MMMk*) oraz kolejne numery podziału *PP=01,02,...* gospodarstwa bazowego.

Przyjęto następujące reguły identyfikacji gospodarstwa bazowego (gospodarstwa, które otrzymuje numer podziału gospodarstwa bazowego *PP*) po podziale tego gospodarstwa:

- gospodarstwo, które pozostaje pod tym samym adresem co gospodarstwo bazowe;
- jeżeli żadne z gospodarstw powstałych w wyniku podziału gospodarstwa bazowego nie pozostało pod dotychczasowym adresem, numer podziału gospodarstwa bazowego otrzymuje gospodarstwo, w którym znajduje się osoba o najniższym numerze porządkowym z poprzedniej rundy badania.

Powyższe reguły są stosowane także w sytuacjach niezmiernie rzadko mogących wystąpić w praktyce, gdy włączone do panelu gospodarstw nowo powstałe gospodarstwo (w wyniku podziału gospodarstw z poprzedniej rundy panelu) składa się z osób pochodzących z dwóch lub więcej gospodarstw należących do panelowej próby gospodarstw.

W sytuacji połączenia dwu lub więcej gospodarstw domowych uczestniczących w pierwszej rundzie badania w jedno gospodarstwo domowe, nowo powstałe gospodarstwo przyjmuje GOSPID gospodarstwa, które nie zmieniło adresu zamieszkania. Pozostałe gospodarstwa zachowują

Identyfikatory Gospodarstwa Domowego z poprzedniej rundy badania i nie podlegają dalszemu badaniu. W sytuacji, gdy nowo powstałe gospodarstwo domowe przeprowadziło się pod nowy adres, przyjmuje ono Identyfikator Gospodarstwa Domowego gospodarstwa domowego, z którego pochodzi aktualna głowa gospodarstwa.

#### 4.7.2. Numery identyfikacyjne osób

Każda z osób należących do panelowej próby gospodarstw w pierwszej rundzie otrzymuje swój unikalny numer identyfikacyjny OSOBID (Identyfikator Osoby). Numer ten jest tworzony poprzez kombinację numeru identyfikacyjnego gospodarstwa GOSPID, w którym ta osoba pojawiła się w badaniu po raz pierwszy oraz dwuznakowego numeru porządkowego osoby w tym gospodarstwie (OO). Ostatecznie unikalny numer identyfikacyjny osoby OSOBID ma postać: GOSPID-OO, czyli: *N-RRRR-MMMk-PP-OO*.

Raz nadany osobie numer identyfikacyjny pozostaje niezmienny we wszystkich rundach panelu, bez względu na to jaki jest aktualnie numer identyfikacyjny gospodarstwa, do którego należy ta osoba. Dla zapewnienia unikalności numeru identyfikacyjnego osoby jej numer porządkowy wewnątrz danego gospodarstwa OO jest niezmienny przez cały okres trwania panelu. Oznacza to, że gdy w gospodarstwie *N-RRRR-MMMk-PP* po raz pierwszy pojawia się w badaniu numer osoby z tego gospodarstwa OO, przenoszony jest on do kolejnej rundy panelu tak długo, jak tylko gospodarstwo *N-RRRR-MMMk-PP* bierze udział w badaniu.

W kolejnej rundzie panelu nowy członek danego badanego gospodarstwa domowego *N-RRRR-MMMk-PP* zostaje dołączony do składu osobowego tego gospodarstwa z kolejnym numerem porządkowym OO, dotąd nie wykorzystywanym. W przypadku nowo włączanych do panelu gospodarstw, które powstały na skutek podziału gospodarstwa *N-RRRR-MMMk-PP* i otrzymały nowe numery identyfikacyjne typu *N-RRRR-MMMk-P'P*, będą utrzymywane numery porządkowe osób OO jakie otrzymały one w gospodarstwach bazowych w kolejnych rundach badania. Powyższe reguły nadawania numerów identyfikacyjnych osobom oznaczają, że w nowo powstałych gospodarstwach z podziału już badanych gospodarstw oraz w gospodarstwach, które wchłonęły inne gospodarstwa z próby panelowej gospodarstw mogą być osoby o różnych GOSPID-ach (pierwszych 11 znakach) i jednakowych ostatnich dwóch znakach. Nadawanie numerów porządkowych członkom nowo włączonego do panelu gospodarstwa domowego (powstałego w wyniku podziału już funkcjonującego w panelu gospodarstwa) rozpoczyna się od przyznania tych samych numerów porządkowych, jakie mieli oni w gospodarstwie bazowym, osobom należącym do panelowej próby osób, które przeniosły się do nowego gospodarstwa.

## 5. Próba oraz uogólnianie wyników badania na populację

### 5.1. Założenia

W przypadku gdy populację generalną stanowią wszystkie gospodarstwa domowe w Polsce, nie ma możliwości sporządzenia ich wykazu na konkretny dzień, czyli ich operatu losowania<sup>2</sup>. W tej sytuacji

---

<sup>2</sup> Operatem losowania nazywamy ponumerowany wykaz jednostek badania lub ich zespołów, zwanych jednostkami losowania (zespoły te muszą być rozłączne), spełniający warunek, że każdej

operatem losowania nie będzie wykaz jednostek badania lecz ich zespołów, tj. obwodów spisowych zawierających mieszkania. Z takiego operatu, według określonego schematu<sup>3</sup>, losowane są obwody spisowe, a w ich ramach wszystkie mieszkania, czyli jednostki badania będą losowane zespołowo<sup>4</sup>. W wylosowanych obwodach spisowych będą następnie losowane mieszkania. W wylosowanych mieszkaniach badane będą wszystkie gospodarstwa domowe oraz ich członkowie.

## 5.2. Schemat losowania próby

Próby na poziomie ogólnokrajowym oraz na poziomie lokalnym były losowane przy wykorzystaniu operatów losowania utworzonych na bazie Urzędowego Rejestru Podziału Terytorialnego Kraju (systemu TERYT). Próba do badania (zarówno na poziomie ogólnokrajowym jak i na poziomie lokalnym) została wylosowana przy zastosowaniu schematu losowania dwustopniowego, polegającego na losowaniu, według wybranego schematu losowania, najpierw jednostek pierwszego stopnia (jps), a następnie w wylosowanych jps-ach wylosowaniu, także według wybranego schematu losowania, jednostek drugiego stopnia (jds).

Schematy losowania na I i II stopniu nie były identyczne. W badaniu został zastosowany kompromisowy wariant alokacji próby mieszkań między województwa, rekomendowany przez Eurostat dla badań realizowanych przez kraje UE, (takich badań jak Europejskie Badanie Dochodów i Warunków Życia - *Survey on Income and Living Conditions-EU-SILC*, por.: Eurostat, Meeting..., 2001), Europejskie Badanie Stanu Zdrowia Ludności - *European Health Survey-EHIS*, czy Kształcenie Dorosłych - *Adult Education Survey-AES*). Jednostki pierwszego stopnia przed losowaniem zostały powarstwowane. Warstwami były województwa (NTS2), a wewnątrz województw jps-y (obwody spisowe) warstwowane według klasy miejscowości. Na obszarach miejskich jps-y były grupowane według wielkości miast. Duże miasta (powyżej 100 tys. mieszkańców) stanowiły samodzielne warstwy. Ponadto, w pięciu największych miastach samodzielnymi warstwami były dzielnice. Na terenach wiejskich warstwami były gminy wiejskie (NTS5) w podregionach (NTS3) lub też gminy wiejskie z kilku sąsiadujących ze sobą powiatów (NTS4). Na I stopniu został zastosowany schemat losowania z prawdopodobieństwami proporcjonalnymi do wartości funkcji cechy dodatkowej, za jaką uznano szacunkową liczbę mieszkań w warstwie, czyli zastosowano losowanie z różnymi prawdopodobieństwami, bez zwracania. Jednostkami losowania pierwszego stopnia (jps) były obwody spisowe.

---

jednostce losowania odpowiada tylko jeden numer w operacie i jednocześnie każdemu numerowi odpowiada tylko jedna jednostka losowania (por. Bracha Cz., *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*, Efekt, Warszawa 1998).

<sup>3</sup> Schematem losowania nazywamy sposób wyboru jednostek, jedna po drugiej, z populacji generalnej ze z góry ustalonym prawdopodobieństwem wyboru dla poszczególnych jednostek w każdym ciągnięciu (tj. ze z góry ustalonym planem losowania).

<sup>4</sup> Jeżeli wszystkie jednostki losowania są jednocześnie jednostkami badania, to mówimy wtedy o losowaniu indywidualnym. Jeżeli warunek ten nie jest spełniony, to losowanie nazywamy zespołowym.



### 5.2.1. Schemat losowania próby na poziomie ogólnopolskim

Liczby mieszkań były losowane w proponowanym wariacie alokacji próby między województwa, według wzoru:

$$m_w = m_o \sqrt{s^2 + (1-s^2) \cdot \left(\frac{1}{\bar{M}} \cdot M_w\right)}, \quad w=1,2,\dots,16,$$

gdzie:

$m_w$  - liczba mieszkań do wylosowania w  $w$ -tym województwie,

$M_w$  - liczba mieszkań w  $w$ -tym województwie,

$\bar{M}$  - przeciętna liczba mieszkań w województwie,

$s$  - parametr względnej ważności oszacowania dla województw względem oszacowań dla Polski,

$m_o$  - parametr, którego wartość ustalana jest iteracyjnie, przy warunku:

$$\sum_{w=1}^{16} m_w = m,$$

gdzie:

$m$  - założona liczebność ogólnopolskiej próby mieszkań.

Alokacja próby mieszkań pomiędzy poszczególne warstwy, w ramach każdego województwa, była proporcjonalna do szacunkowej liczby mieszkań w tych warstwach w operacie losowania (zapisanych w systemie TERYT), a mianowicie:

$$m_{wh} \approx m_w \cdot \frac{M_{wh}}{M_w}, \quad h=1,2,\dots,W_w, \quad w=1,2,\dots,16,$$

gdzie:

$M_{wh}$  - liczba mieszkań w  $h$ -tej warstwie  $w$ -tego województwa,

$W_w$  - liczba warstw w  $w$ -tym województwie.

Liczba obwodów spisowych losowana na I stopniu w poszczególnych warstwach w ramach województw była proporcjonalna do liczby mieszkań i obliczana jest według wzoru:

$$n_{wh} = \frac{m_{wh}}{\bar{m}_{wh}}, \quad h=1,2,\dots,W_w, \quad w=1,2,\dots,16,$$

gdzie:

$m_{wh}$  - liczba mieszkań do wylosowania w  $h$ -tej warstwie  $w$ -tego województwa,

$\bar{m}_{wh}$  - liczba mieszkań do wylosowania w pojedynczym obwodzie spisowym w  $h$ -tej warstwie  $w$ -tego województwa.

Próba I stopnia (obwody spisowe) była losowana oddzielnie w każdej warstwie każdego z województw, zgodnie z schematem losowania Hartley'a i Rao (1962). Losowanie według tego

schematu polega na systematycznym doborze jednostek losowania (w UDE - obwodów spisowych) po ich uprzednim uporządkowaniu losowym. Na wstępie obwody spisowe w danej warstwie zostały uporządkowane losowo. Następnie utworzono ciąg  $\{S_{who}\}$  zawierający skumulowane liczby mieszkań w obwodach spisowych danej warstwy:

$$S_{who} = S_{wh,o-1} + M_{who}, \quad o=1,2,\dots,N_{wh}, \quad h=1,2,\dots,W_w, \quad w=1,2,\dots,16,$$

gdzie:

$M_{who}$  - liczba mieszkań w  $o$ -tym obwodzie  $h$ -tej warstwy  $w$ -tego województwa,

$N_{wh}$  - liczba obwodów spisowych w  $h$ -tej warstwie  $w$ -tego województwa,

przy czym:

$$\begin{aligned} S_{who} &= 0 && \text{dla } o = 0, \\ S_{who} &= M_{wh} && \text{dla } o = N_{wh}. \end{aligned}$$

W kolejnym kroku obliczono interwał losowania (liczbę rzeczywistą) według wzoru:

$$q_{wh} = \frac{M_{wh}}{n_{wh}}, \quad h=1,2,\dots,W_h, \quad w=1,2,\dots,16,$$

gdzie:

$q_{wh}$  - interwał losowania w  $h$ -tej warstwie  $w$ -tego województwa.

W kolejnym kroku losowano początek losowy  $R_{wh}$  (liczbę rzeczywistą) z przedziału  $(0; q_{wh})$ . Na podstawie wartości  $q_{wh}$  i  $R_{wh}$  tworzymy ciąg liczbowy  $\{a_{whi}\}$  według wzoru:

$$a_{whi} = R_{wh} + (i-1)q_{wh}, \quad i=1,2,\dots,m_{wh}, \quad h=1,2,\dots,W_h, \quad w=1,2,\dots,16.$$

Do próby zostały wybrane w danej warstwie obwody spisowe, dla których spełniony był warunek:

$$S_{wh,o-1} < a_{whi} \leq S_{who}, \quad o=1,2,\dots,N_{wh}, \quad h=1,2,\dots,W_w, \quad w=1,2,\dots,16.$$

Na II stopniu zastosowano schemat losowania prostego bez zwracania. Jednostkami losowania drugiego stopnia (jds) w wylosowanych na I stopniu obwodach były mieszkania. Ostatecznie badaniu podlegały, w wylosowanych obwodach, wszystkie gospodarstwa domowe i osoby w wieku 15-65 lat (wywiady indywidualne) zamieszkałe w wylosowanych mieszkaniach.

W obliczeniach wskaźników na poziomie ogólnopolskim nieproporcjonalna alokacja próby mieszkań między województwami została zrekompensowana przy konstrukcji systemu wag (wagi te są odpowiednio mniejsze dla terenów nadreprezentowanych i odpowiednio większe dla terenów niedoreprezentowanych).

### 5.2.2. Schemat losowania próby na szczeblu lokalnym

Punktem wyjścia dla doboru próby na szczeblu lokalnym była klasyfikacja każdego z typów powiatów (osobno ziemskich i osobno grodzkich) do czterech homogenicznych grup pod względem poziomu rozwoju społeczno-ekonomicznego. Ze względu na cel projektu wskaźniki opisujące rozwój

społeczno-ekonomiczny powiatów dotyczyły czterech obszarów, tj. poziomu rozwoju ekonomicznego i zamożności, rynku pracy, kapitału ludzkiego i sytuacji demograficznej oraz edukacji. Klasyfikacja powiatów została przeprowadzona za pomocą jednej z metod statystycznej analizy porównawczej (Panek, 2009). Ponadto został dokonany podział województw na dwie grupy ze względu na ich poziom rozwoju ekonomicznego, mierzony wielkością PKB na mieszkańca. Do badań na szczeblu lokalnym ostatecznie wybrano w sposób celowy po jednym województwie z każdej z grup województw o podobnym poziomie rozwoju społeczno-ekonomicznego. W każdym z wytypowanych województw wybrano losowo, w każdej z wyodrębnionych w początkowym etapie grup powiatów ziemskich oraz powiatów grodzkich, po jednym powiecie reprezentancie. (Tabela 4).

**Tabela 4.** Powiaty reprezentanty w grupie powiatów ziemskich

Województwa	Powiaty ziemskie			
	Grupa 1	Grupa 2	Grupa 3	Grupa 4
Lubuskie			0809 zielonogórski	0806 strzelecko-drezdenecki
Małopolskie	1216 tarnowski	1203 chrzanowski	1219 wielicki	1205 gorlicki
Podkarpackie	1807 krośnieński	1817 sanocki		

## 6. Ustalenie minimalnej wielkości próby

Przy ustalaniu minimalnej liczebności próby bierze się pod uwagę następujące czynniki:

- postać parametru, który będzie szacowany na jej podstawie dla populacji generalnej;
- liczebność populacji generalnej;
- stopień precyzji jaki należy osiągnąć przy estymacji parametru;
- schemat losowania próby oraz jego efektywność w stosunku do schematu losowania prostego ze zwracaniem.

W projekcie minimalna liczebność prób powinna zapewnić odpowiednią precyzję szacunku zarówno na poziomie krajowym i wojewódzkim (badanie ogólnokrajowe) jak i na poziomie powiatowym (badanie lokalne).

Dla szacunku minimalnej liczebności próby (liczby mieszkań), przy założeniu, że szacowanym parametrem jest frakcja elementów z wyróżnioną cechą oraz przyjętym schematem losowania próby, stosuje się wzór o następującej postaci:

$$n_{\min} = \frac{def^2(1-\hat{p})\left(1-\frac{n}{N}\right)}{\hat{p} \cdot RSE^2},$$

gdzie:

$n_{\min}$  - minimalna liczebność próby,

*deft* - wskaźnik efektywności planu losowania próby,

$\hat{p}$  - przybliżona wartość parametru w populacji generalnej znana z zewnętrznych źródeł.

Podstawą do wykorzystania tej reguły jest założenie, że w projekcie najczęściej szacowanym parametrem w populacji generalnej jest funkcja elementów z wyróżnioną cechą (np. odsetki gospodarstw domowych osób charakteryzujących się określoną właściwością, np. stopa bezrobocia). Przy szacunku frakcji cech wielowariantowych przyjmuje się na ogół konserwatywne założenie, że  $\hat{p} = 0,08$ . Zgodnie z ogólnie obowiązującymi standardami statystycznymi, za dostateczną precyzję uznaje się poziom względnego błędu szacunku *RSE* poniżej 15% (Zasępa, 1991). Z kolei efektywność schematu losowania w stosunku do schematu prostego, mierzona stosunkiem względnych błędów szacunku uzyskiwanych w estymacji przy tych schematach (tzw. wskaźnik *deft* - efektywności planu losowania próby), w prowadzonych w Polsce badaniach (EU-SILC, EHIS) dla cech demograficznych przyjmował wartości nieznacznie mniejsze od 1, a dla cech społeczno-ekonomicznych nieznacznie większe od 1. Przyjmując w powyższym wzorze, że przeciętna liczba mieszkań w województwie wynosi 838876 mieszkań, przeciętna liczba mieszkań w powiecie (po wyłączeniu 5 powiatów stanowiących miasta wydzielone, które nie będą badane na poziomie lokalnym ze względu na ich nietypowość oraz dysponowanie licznymi próbami w tych powiatach w badaniu na poziomie ogólnokrajowym) to 30530 mieszkań, zaś wartość *deft* równa się 1,2 (założenie konserwatywne) oraz wartość względnego błędu szacunku *RSE* wynosi 0,15 (wpływ współczynnika korygującego  $(1-n/N)$  na liczebność próby dla badania na poziomie województwa przy dużej liczebności populacji z jaką mamy do czynienia na poziomie wojewódzkim jest zawsze nieznaczny), uzyskano następującą przeciętną minimalną liczebność próby na poziomie wojewódzkim:

$$n_{\min,w} = \frac{1,2(1-0,08)\left(1 - \frac{1562}{838876}\right)}{0,08 \cdot (0,15)^2} = 612$$

Ze względu na niemożność przewidzenia, czy braki odpowiedzi w drugiej rundzie panelu nie przekroczą 30 procent próby (warunek wysoce "wyśrubowany"), za rozsądną minimalną liczebność próby w badaniu na poziomie ogólnopolskim zaproponowano 24 000 gospodarstw domowych. Ze względu na fakt, że średnia liczebność populacji w województwie jest 2,75 razy wyższa niż średnia liczebność próby w powiecie, liczebność próby w powiecie powinna wynosić przeciętnie 545 gospodarstw, a w skali całego badania na poziomie lokalnym (16 powiatów) w przybliżeniu 8900 gospodarstw.

## 7. Próby rezerwowe

Założono, że w wypadku braku możliwości nawiązania kontaktu z respondentem lub jego odmowy udziału w badaniu należy skorzystać z prób rezerwowych, przygotowanych wcześniej i wylosowanych wraz z próbą podstawową. Założono, że minimalna liczba prób rezerwowych wynosi 3, ale ich liczba była dostosowana do wyników nawiązywania kontaktów w terenie.

Każda próba rezerwowa stanowi 100% próby podstawowej, czyli do pierwszej rundy wylosowano od razu 3 próby rezerwowe obok próby podstawowej, wszystkie o identycznej strukturze. Założono, że w drugiej rundzie dopełnianie próby podstawowej spowoduje konieczność dodatkowego kontaktu z około 30% dodatkowymi gospodarstwami (wielkość orientacyjna zależna od skali erozji panelu). Losowanie próby rezerwowej do drugiej rundy badania (próba uzupełniająca) jest dokonane

jednocześnie z próbą pierwotną i adresy są losowane w tych samych obwodach spisowych. Adresy rezerwowe odnoszą się do tych samych obwodów. Próby rezerwowe gwarantują, że liczebność próby w każdej rundzie odpowiada liczebności próby wyjściowej z rundy pierwszej, co schematycznie ilustruje tabela 5.

**Tabela 5.** Pozostawanie gospodarstw domowych w badaniu przekrojowym

<b>Procent gospodarstw domowych pozostających w badaniu przekrojowym</b>	
<b>długość panelu</b>	
1-runda	2-runda
100	100

## 8. Wagi

System wag musi być konstruowany dla każdego etapu badania zarówno dla analiz przekrojowych jak i longitudinalnych. Wagi dla pierwszej rundy panelu (dla początkowej próby) mają na celu przywrócenie w trakcie obliczeń pierwotnej struktury próby, zniekształconej odmowami uczestnictwa w badaniu (odmowami badanych gospodarstw i ich członków). Ważenie w pierwszym etapie badania może mieć także na celu skorygowanie rozkładów cech z próby (tak gospodarstw 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.

W drugiej rundzie panelu ważenie ma na celu korygowanie zniekształceń próby powstających na skutek ubytku badanych jednostek (gospodarstw domowych i osób) powodowanych odmowami i utratą kontaktu z nimi oraz włączaniem do próby nowo tworzonych gospodarstw przez osoby należące do panelowej próby osób oraz zmianami struktury badanych gospodarstw. Zmiany powodowane wymieraniem jednostek nie powinny być korygowane, gdyż ubytki tego typu są reprezentatywne dla populacji.

### 8.1. Wagi przekrojowe

Wagi przekrojowe są obliczane w celu szacunku parametrów dla próby wylosowanej w 2013 r. oraz podpróbki, która została włączona do badania w drugiej rundzie. Przy szacunku tych wag uwzględniono:

- prawdopodobieństwa wyboru do próby mieszkań, gospodarstw domowych i osób;
- poziom kompletności badania;
- strukturę demograficzną populacji gospodarstw domowych i osób w populacji generalnej na podstawie danych pochodzących z zewnętrznych źródeł (NSP 2011).

W ramach badania obliczono następujące podstawowe wagi przekrojowe:

- dla gospodarstw domowych;
- dla osób, członków gospodarstwa domowego;
- dla osób w wieku 15-65 lat (osób, z którymi przeprowadzany jest wywiad indywidualny).

Punktem wyjścia do obliczeń wag przekrojowych było wyznaczenie wag początkowych dla wszystkich gospodarstw domowych wylosowanych po raz pierwszy do badania. Oznacza to, że wagi początkowe obliczono dla wszystkich gospodarstw domowych wylosowanych do próby w 2013 r. oraz nowo wprowadzanych do badania podróbek w drugiej rundzie badania (2014 r. podróbka o numerze 2, por. tabela 1) wchodzących w skład panelowej próby gospodarstw domowych.

Waga początkowa dla gospodarstwa domowego wynika z przyjętego planu losowania próby i jest równa odwrotności prawdopodobieństwa wyboru mieszkania, w którym zamieszkuje badane gospodarstwo domowe, w  $h$ -tej warstwie:

$$wg_{whj\epsilon i}^{pocz} = \frac{1}{\pi_{whi}}, \quad j=1,2,\dots,n; \quad h=1,2,\dots,W_w; \quad w=1,2,\dots,16,$$

gdzie:

$\pi_{whi}$  - prawdopodobieństwo wylosowania  $i$ -tego mieszkania, w którym zamieszkuje badane gospodarstwo domowe, w  $h$ -tej warstwie  $w$ -tego województwa,

przy czym:

$$\pi_{whi} = \frac{m_{wh} \bar{m}_{wh}}{M_{wh}}, \quad j=1,2,\dots,n; \quad h=1,2,\dots,W_w; \quad w=1,2,\dots,16,$$

gdzie:

$M_{wh}$  - liczba mieszkań w  $h$ -tej warstwie  $w$ -tego województwa,

$m_{wh}$  - liczba mieszkań do wylosowania w  $h$ -tej warstwie  $w$ -tego województwa,

$\bar{m}_{wh}$  - liczba mieszkań do wylosowania w pojedynczym obwodzie spisowym w  $h$ -tej warstwie  $w$ -tego województwa.

W kolejnym kroku waga  $wg_{whj\epsilon i}^{pocz}$  jest korygowana ze względu na kompletność badania, na poziomie gospodarstwa domowego, odpowiednim wskaźnikiem w celu zmniejszenia obciążenia estymatorów. Wskaźnik ten uwzględnia niemożność skontaktowania się z gospodarstwami domowymi w wylosowanym mieszkaniu oraz niemożność przeprowadzenia wywiadu w gospodarstwach, z którymi nastąpił kontakt (odmowy). Korekta wag początkowych w badaniu przeprowadzonym w 2013 r. będzie dotyczyła wszystkich jednostek składających się na próbę badaną w tym roku. Począwszy od roku 2014 (czyli od drugiej rundy badania) korekta ta dotyczy wyłącznie nowo wprowadzanych podróbek.

Rozwiązaniem idealnym byłaby znajomość prawdopodobieństw kompletności odpowiedzi dla każdego z ankietowanych gospodarstw, a następnie przeważenie ich przez odwrotności tych prawdopodobieństw. Niestety wielkości te nie są znane i trzeba je oszacować. W tym celu próba gospodarstw domowych została podzielona na grupy. W prowadzonym badaniu grupowanie gospodarstw zostało dokonane ze względu na klasę miejscowości zamieszkania gospodarstwa

domowego, przy czym wyróżniono 6 takich klas. Założono, że prawdopodobieństwo kompletności odpowiedzi w każdej z klas jest stałe. Innymi słowy zaobserwowany wewnątrz danej klasy wskaźnik kompletności odpowiedzi stanowi oszacowanie wskaźnika kompletności odpowiedzi dla każdego gospodarstwa należącego do tej klasy.

Wskaźniki kompletności dla kwestionariusza gospodarstwa domowego, dla każdej z wyróżnionych klas miejscowości, zostały obliczone według wzoru:

$$KG_s = KG1_s \cdot KG2_s, \quad s=1,2,\dots,6,$$

gdzie:

$KG1_s$  - wskaźnik kontaktu z wylosowanym gospodarstwem domowym z  $s$ -tej klasie miejscowości,

$KG2_s$  - wskaźnik przeprowadzenia wywiadu z gospodarstwem domowym w  $s$ -tej klasie miejscowości.

Pierwszy ze wskaźników ( $KG1_s$ ) jest ilorazem liczby mieszkań, w których mieszkają gospodarstwa domowe, z którymi ankieterom udało się skontaktować, z liczbą faktycznie istniejących mieszkań. Przy obliczaniu wskaźnika nie są brane pod uwagę mieszkania zlikwidowane, zamienione w budynki niemieszkalne czy też błędne adresy. Drugi ze wskaźników ( $KG2_s$ ) stanowi frakcję gospodarstw, z którymi przeprowadzono wywiad. Przy szacowaniu wartości wskaźników  $KG1_s$  i  $KG2_s$  zostaną zastosowane wagi początkowe ( $wg_{whj}^{pocz}$ ).

Korekta wag początkowych gospodarstw domowych ze względu na kompletność badania przebiegała według wzoru:

$$wg_{wh,j \in s}^{skor} = \frac{wg_{wh,j \in s}^{pocz}}{KG_s}, \quad j=1,2,\dots,n, \quad s=1,2,\dots,z,$$

gdzie:

$wg_{wh,j \in s}^{skor}$  - waga  $j$ -tego gospodarstwa zamieszkującego  $s$ -tą klasę miejscowości zamieszkania w  $h$ -tej warstwie  $w$ -tego województwa.

Kolejnym krokiem w obliczaniu wag przekrojowych był szacunek, dla każdej podpróby, tzw. wag bazowych dla danej rundy (roku) badania ( $t$ ). Pozwalają one uogólniać wyniki zarówno dla gospodarstw domowych jak i ich członków na badaną populację w danej rundzie.

Wagi skorygowane poddawane są tzw. procedurze zintegrowanej kalibracji (Description..., 2010, Estevao, 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 kalibracji zostały wykorzystane następujące zmienne:

- na poziomie gospodarstwa domowego: wielkość gospodarstwa domowego (wyróżniono 4 kategorie wielkości: 1-osobowe, 2-osobowe, 3-osobowe, z 4 osobami i więcej), województwo zamieszkania, typ obszaru zamieszkania (obszar wiejski i obszar miejski);
- na poziomie osób: płeć, grupa wiekowa (wyróżnionych zostanie 14 grup wieku: poniżej 15 lat, dwanaście pięcioletnich grup, 75 lat i więcej).

Informacje dotyczące zmiennych kalibracji pochodziły z Narodowego Spisu Powszechnego Ludności i Mieszkań 2011 oraz z bieżących szacunków demograficznych.

Wartości zmiennych dotyczące osób są w pierwszym kroku agregowane w ramach poszczególnych gospodarstw domowych poprzez obliczanie wartości ogółem tych zmiennych w ramach gospodarstw (jak na przykład liczba kobiet/mężczyzn w określonej grupie wiekowej w gospodarstwie). Następnie przeprowadzona jest kalibracja na poziomie gospodarstwa, wykorzystująca zmienne dotyczące gospodarstw (na przykład liczba gospodarstw na obszarze wiejskim w danym województwie) oraz zmienne dotyczące osób w formie zagregowanej. Zaletą tej techniki jest zapewnienie zgodności pomiędzy estymacją dotyczącą gospodarstw oraz estymacją dotyczącą osób, gdyż wszyscy członkowie gospodarstw domowych (osoby) otrzymują te same wagi przekrojowe co gospodarstwa, do których należą.

Procedura kalibracji wag  $wg_{whj}^{skor}$  dla gospodarstw domowych polega na takiej ich korekcie aby wyniki dla badanej populacji otrzymywane na podstawie próby przy wykorzystaniu wag skalibrowanych  $wg_j^{skalib}$  jak najlepiej odtwarzały strukturę badanej populacji opisaną zmiennymi kalibracji, których wartości pochodzą z zewnętrznych źródeł. Jednocześnie ilorazy tych wag po kalibracji i przed kalibracją powinny być jak najbliższe jedności (korekta wag za pomocą procedury kalibracji powinna być jak najmniejsza).

Wagi skalibrowane poddawane są następnie procedurze obcinania wag ekstremalnych (Description..., 2010, Elliott M., Little, 2000). Zbyt duże zróżnicowanie wag jest niekorzystne dla precyzji estymacji, gdyż zwiększa wariancję estymatorów. W efekcie stosowania procedury obcinania wag ekstremalnych otrzymano wagę ostateczną dla gospodarstw domowych ( $w_j^{ost}$ ) oraz ich członków ( $w_l^{ost}$ ) dla próby podlegającej badaniu po raz pierwszy w 2013 r.

Wagi ostateczne w 2013 r. są jednocześnie wagami bazowymi. Od drugiej rundy badania postępuje się analogicznie, kalibrując i obcinając wagi ekstremalne. W 2014 r. wagi ostateczne dla podpróbek 1 i 2 są przemnażane przez odwrotności udziału ich liczebności w liczebności obu podróbek łącznie i stają się wagami bazowymi. Dla analiz przekrojowych w 2014 r. połączono obie podróbki. Wagi w każdej z podróbek są dzielone przez odwrotności udziału ich liczebności w liczebności obu podróbek łącznie i poddawane algorytmowi zintegrowanej kalibracji oraz obcinania wag ekstremalnych.

W przypadku uogólniania wyników, dotyczących członków gospodarstw domowych podlegających wywiadowi indywidualnemu (w naszym badaniu osób w wieku 15-65 lat), na całą populację tych osób konieczna była korekta ich wag ze względu na niekompletność badania na poziomie osób (niekompletność wywiadów indywidualnych). Miała ona na celu doprowadzenie do zgodności rozkładów wybranych cech osób podlegających wywiadowi indywidualnemu w badanej próbie i w całej populacji.

Wskaźniki korygujące są stosowane dla każdej z grup osób wyróżnionych ze względu na wiek i płeć (identycznych jak podczas korygowania wag początkowych gospodarstw domowych), oddzielnie dla poszczególnych województw, a w ich ramach w podziale na miasto i wieś. Korekta ta przebiega według wzoru:

$$w_{wh,l \in s}^{ost} = KO_s \cdot w_{wh,l \in s}^{skalib}, \quad l=1,2,\dots,z', \quad h=1,2,\dots,W_w; \quad w=1,2,\dots,16.$$

gdzie:



$WO_{wh,l \in S}^{ost}$ ,  $WO_{wh,l \in S}$  - waga  $l'$ -tej osoby podlegającej wywiadowi indywidualnemu należącej do  $s$ -tej grupy osób.

Wagi bazowe są punktem wyjścia zarówno do szacunku ostatecznych wag przekrojowych dla danej rundy badania jak i dla obliczeń wag bazowych poszczególnych podpróbek badanych ponownie w kolejnej rundzie badania.

Dla nowo wylosowanej podróbki w 2014 r. przeprowadzamy identyczną jak w 2013 r procedurę obliczania wag bazowych dla gospodarstw domowych i osób.

Dla podpróbek, które są badane po raz drugi w 2014 r. (podpróbka 2) punktem wyjścia do obliczeń wag bazowych dla gospodarstw domowych i osób w tych gospodarstwach jest szacunek wag dla osób. Obliczenia są przeprowadzone na panelowej próbie osób będących członkami podpróbek panelowych gospodarstw domowych, które to osoby zbadano w 2013 r. i powinny zostać zbadane w 2014 r.

Wagi bazowe dla osób z roku 2012 korygowane są odpowiednio współczynnikami ze względu na braki kompletności odpowiedzi oraz „wypadanie” gospodarstw domowych i osób poza badane podpróbki. Współczynniki są obliczane oddzielnie dla każdej klasy miejscowości zamieszkania (dla podpróbki 1) według wzoru:

$$KO_s(2) = \frac{z_s^p(1) - z_s^{np}(1,2)}{z_s^p(2)},$$

gdzie:

$z_s^p(1) = z_s^p(2)$  - liczba osób należących do panelowej próby osób, w  $s$ -tej klasie miejscowości w podpróbie gospodarstw domowych badanej odpowiednio po raz 1-szy i 2-gi (w 2013 r. i 2014 r.),

$z_s^{np}(1,2)$  - liczba osób z panelowej próby osób, w  $s$ -tej klasie miejscowości zamieszkania, które były zbadane w 1-szej rundzie badania i wypadły z badania w 2-giej rundzie badania (wypadły z badania ich gospodarstwa domowe lub też osoby).

Powyższy wskaźnik korygujący jest odwrotnością oszacowania prawdopodobieństwa, że dana osoba z  $s$ -tej klasy miejscowości zamieszkania zbadana w 1-szej rundzie badania zostanie zbadana ponownie w 2-giej rundzie badania. W kolejnym etapie obliczeń wag bazowych dla osób w 2014 r. następuje przypisanie wag matek, które należą do panelowej próby osób, nowonarodzonym dzieciom oraz wag zerowych osobom, które nie należały do gospodarstw domowych stanowiących próbę panelową gospodarstw w 2013 r. (współmieszkańcom), a znalazły się w tych gospodarstwach w 2014 r.

Następnie w ramach każdego badanego gospodarstwa oblicza się średnią arytmetyczną z wag osób należących do tego gospodarstwa, która zostaje przypisana wszystkim członkom tego gospodarstwa. Otrzymane w ten sposób wagi są poddane procedurze kalibracji, a następnie procedurze obcinania wag ekstremalnych. W ostatnim kroku dokonywana jest korekta wag osób podlegających wywiadowi indywidualnemu ze względu na niekompletność badania na poziomie osób.

Szacunki parametrów z populacji generalnej dla każdej rundy badania wykonywane są na wszystkich połączonych podpróbkach badanych w danej rundzie. W obliczeniach stosowane są wagi przekrojowe dla danej rundy badania. Zbiory tych wag ze wszystkich podpróbek łącznie poddawane są wcześniej opisanym procedurom kalibracji i obcinania wag ekstremalnych.

## 8.2. Wagi longitudinalne

Każda analiza longitudinalna musi dotyczyć konkretnego przedziału czasowego, od rundy w roku  $t-2$  do rundy w roku  $t$  i obejmować osoby, które należałyby do panelowej próby osób w czasie tego całego przedziału czasowego. To samo dotyczy wag odpowiadających osobom.

Stają się one wagami longitudinalnymi tych osób w analizowanych przedziale czasowym oraz wagami longitudinalnymi gospodarstw domowych, których członkami są te osoby.

## 9. Metoda oceny precyzji estymacji

Procedura estymacji parametrów prowadzi do uzyskania złożonych estymatorów zaliczających się do estymatorów ilorazowych. Błędy standardowe takich estymatorów, będące miarą precyzji ich szacunku, nie dają się oszacować klasycznymi metodami. W projekcie estymacja błędów standardowych zostanie oparta na jednej z metod bootstrap (Kordos, Zięba-Pietrzak, 2010).

Dla zbadania efektywności zastosowanego w badaniu schematu losowania próby w stosunku do losowania prostego, stanowiącego punkt odniesienia do oceny precyzji szacunku parametrów, zostaną oszacowane wskaźniki efektywności planu losowania próby o postaci (Kish, 1995):

$$defr(\hat{\theta}) = \frac{ASE(\hat{\theta})}{ASE(\hat{\theta}_p)}$$

gdzie:

$ASE(\hat{\theta}_p)$  - absolutny błąd standardowy szacunku parametru  $\hat{\theta}$  w przypadku zastosowania schematu losowania prostego.

## Literatura

*Description of target variables: Cross-sectionnal and longitudinal*, EU-SILC 065/2010, European Commission, Eurostat, Luxembourg 2010, str. 30-42

Elliott M., Little R., *Model based alternatives to trimming survey weights*, "Journal of Official Statistics" 2000, Vol. 16, No. 3, str. 191-209.

Estevao V., Särndal C-E, *Survey estimates by calibration and complex auxiliary information*, "Internal Statistical Review" 2006, Vol. 74, No. 2, str. 127-147.

Eurostat, *Meeting of the Task Force Statistics on Income and Living Conditions (EU-SILC)*, 24-25 September 2001, Luxembourg.

GUS, *Metodologia badania budżetów gospodarstw domowych*, Zeszyty Metodologiczne, Warszawa, 2011.

Hartley O. H. I Rao J. N., *Sampling with unequal probabilities and without replacement*, "Annals of Mathematical Statistics" 1962, vol. 33, str. 350-374.

Kish L., *Methods for design effect*, "Journal of Official Statistics" 1995, Vol. 11, No 1 , str. 55-77.

Kordos J., Zięba-Pietrzak A., *Development of standard error estimation methods in complex household sample survey in Poland*, "Statistics in Transition" 2010, Vol. 11, No 2, str. 231-252.

Panek T., *Statystyczne metody wielowymiarowej analizy porównawczej*, Oficyna Wydawnicza SGH, Warszawa 2009.

Zasępa R., *Zarys metody reprezentacyjnej*, Biblioteka Wiadomości Statystycznych, t. 39, GUS, Warszawa 1991.