

Державний комітет статистики України

ЗАТВЕРДЖЕНО

Наказ Держкомстату

31.10.2011 №278

МЕТОДИКА

**обробки результатів державного вибіркового
обстеження умов життя домогосподарств**

Київ - 2011

Методика обробки результатів державного вибіркового обстеження умов життя домогосподарств (далі – Методика) визначає методичні засади основних етапів обробки цього обстеження на регіональному та центральному рівнях (уведення до ПЕОМ первинних даних обстеження, контроль та імпутація даних, формування файлів аналізу, розрахунок статистичних ваг і підготовка аналітичних таблиць для публікацій). У документі також наведені концептуальні підходи до обробки результатів обстеження.

Методика підготовлена експертом проекту Світового банку "Розвиток системи державної статистики для моніторингу соціально-економічних перетворень" Лисою О. за участю директора департаменту обстежень домогосподарств Осипової І.І. та начальника відділу підготовки та поширення даних обстеження умов життя домогосподарств департаменту обстежень домогосподарств Рокитько Л.Л.

Методика призначена для використання працівниками органів державної статистики, міністерств, інших органів виконавчої влади, наукових організацій, навчальних закладів, іншими користувачами статистичної інформації, яка розробляється за підсумками обстеження умов життя домогосподарств.

ЗМІСТ

Перелік скорочень	4
Визначення термінів	4
Передмова	6
1. Загальні положення	7
2. Обробка результатів обстеження на регіональному рівні	9
2.1. Візуальний контроль заповнених первинних статистичних формулярів	9
2.2. Уведення первинних даних обстеження до ПЕОМ	10
2.3. Контроль первинних даних обстеження	14
3. Обробка результатів обстеження на центральному рівні	18
3.1. Функції та завдання, що виконуються у Головному міжрегіональному управлінні статистики у м. Києві	18
3.1.1. Формування єдиної бази даних обстеження на основі регіональних фрагментів	19
3.1.2. Конвертація даних обстеження у формат SPSS, їх контроль і очищення	19
3.1.3. Імпутація відсутніх даних обстеження	23
3.1.4. Файли даних та їх зберігання	24
3.2. Функції та завдання, що виконуються у Держкомстаті	25
3.2.1. Логічний контроль і узгодження файлів даних	25
3.2.2. Формування файлів аналізу даних	29
3.2.3. Формування системи статистичних ваг	32
3.2.4. Аналіз зіставності/порівнянності та узгодженості/послідовності результатів обстеження	35
3.2.5. Розрахунки характеристик надійності оцінок показників	35
3.2.6. Підготовка інформаційно-аналітичних матеріалів за результатами обстеження	37
Додатки	39

ПЕРЕЛІК СКОРОЧЕНЬ

БД – база даних

ГМУС у м. Києві – Головне міжрегіональне управління статистики у м. Києві

КІСЦ – класифікація індивідуального споживання за цілями

КОАТУУ – класифікатор одиниць адміністративно-територіального устрою України

КЕОІ – комплекс електронної обробки інформації

НЗВ – не знаю відповіді

ОУЖД – обстеження умов життя домогосподарств

ПЕОМ – персональна електронно-обчислювальна машина

ПТОВ – первинна територіальна одиниця вибірки

ВТОВ – вторинна територіальна одиниця вибірки

SPSS (Statistical Package for Social Science) – Статистичний пакет для соціальних наук

ВИЗНАЧЕННЯ ТЕРМІНІВ

Вибіркове обстеження населення (домогосподарств) – науково обґрунтований вид несуттєвого спостереження, при якому обстежуються не всі одиниці сукупності, а лише певним чином відібрана їх частина, а одиницею спостереження є особа чи домогосподарство.

Вибірковий метод спостереження – науково обґрунтована система правил відбору одиниць і способів характеристики сукупності досліджуваних одиниць, що вивчають. Вибірковий метод спостереження дає змогу розповсюдити висновки, отримані на основі вивчення частини сукупності (вибірки), на всю сукупність (генеральну).

Вторинна територіальна одиниця вибірки (ВТОВ) – територіальна одиниця, яка відібрана на другому ступені формування територіальної вибірки при застосуванні багатоступеневого територіального відбору.

Генеральна сукупність – множина одиниць, характеристику яких необхідно дослідити і щодо яких буде зроблено узагальнення за результатами вибіркового обстеження.

Глобальний контроль – система контролю, що складається з усіх груп перевірок фрагментарного контролю, перехресного контролю розділів первинного документа, контролю повноти заповнення документа, перевірок узгодженості документа з іншими видами інструментарію.

Дизайн (план) вибірки – організаційно-логічна модель структури вибіркової сукупності та принципи її формування. Якістю плану вибірки значною мірою визначається репрезентативність вибіркової сукупності, отриманої за визначеними у плані правилами.

Дизайн-ефект (deff) – міра відносної статистичної ефективності певного плану вибірки порівняно з простою випадковою вибіркою (відношення дисперсії вибірових оцінок показника для плану вибірки, реалізованого в обстеженні, до дисперсії оцінки цього показника, отриманої за умови використання процедури простого випадкового відбору).

Дисперсія вибірових оцінок – середній квадрат відхилень оцінок показника за всіма можливими вибірками цього дизайну (реплікаціями) від середнього значення оцінки за всіма вибірками.

Домогосподарство – сукупність осіб, які спільно проживають в одному житловому приміщенні або його частині, забезпечують себе всім необхідним для життя, ведуть спільне господарство, повністю або частково об'єднують та витрачають кошти. Ці особи можуть перебувати в родинних стосунках або стосунках свояцтва, не перебувати в будь-яких із цих стосунків або бути і в тих, і в інших стосунках. Домогосподарство може складатися з однієї особи (стаття 1 Закону України “Про Всеукраїнський перепис населення”).

Імпутація – технологічна операція, що полягає у визначенні та введенні до обстежуваної сукупності значень для специфічних даних, щодо яких інформація відсутня або не може бути використана.

Калібрація системи статистичних ваг – трансформація (коригування) системи ваг для одиниць спостереження, які взяли участь у вибірковому обстеженні, шляхом мінімізації функції відстані між калібраційними вагами та вагами, що коригуються, за умови, що калібраційні ваги дають точні оцінки зовнішніх змінних, відомих для генеральної сукупності або її частин.

Коефіцієнт варіації (відносна стандартна похибка) дає уявлення про ступінь статистичної однорідності результатів при багаторазовому застосуванні процедур вибіркового обстеження однакового дизайну вибірки (реплікацій). Визначається як відношення величини стандартної похибки вибірки до оцінки показника у відсотках, чим менша величина коефіцієнта варіації, тим менше результати за різними реплікаціями відрізняються один від одного, тобто дані більш надійні.

Макрос (макрокоманда) – спеціальний програмний вираз (програма), написаний мовою синтаксису SPSS з метою полегшення обробки даних у середовищі SPSS Statistics. Макрос складається з тіла (власне програма) і пускової команди (команда в SPSS-синтаксисі, яка складається з імені конкретного макросу і параметрів, значення яких повинен задати користувач). Мета функціонування макросу полягає в тому, що за допомогою передачі на вхід до макросу різних значень параметрів можна виконувати записані в програмі макросу дії з різними об'єктами та структурами даних.

Метод імпутації "hot-deck" – метод імпутації, який передбачає вибір даних, що імпутуються, для кожного пропущеного значення з урахуванням розподілу ознаки і ступеня схожості елементів вибірки за деяким набором ознак (набором класифікаційних змінних).

Оцінка показника – значення показника, отримане за неповними даними щодо сукупності, яка досліджується. Може бути отримано шляхом розповсюдження даних вибіркової сукупності на генеральну, застосування моделей, визначення значень експертами тощо.

Первинна територіальна одиниця вибірки (ПТОВ) – територіальна одиниця, яка відібрана на першому ступені формування територіальної вибірки при застосуванні багатоступеневого територіального відбору.

Похибка вибірки – частина розбіжності між істинним значенням статистичного показника та його значенням, розрахованим за вибіркою, що зумовлена застосуванням саме несущільного характеру обстеження.

Середньоквадратична похибка – величина, що характеризує надійність оцінювання показників за результатами вибірових обстежень за умов зміщення оцінок. Середньоквадратична похибка визначається як сума дисперсії вибірових оцінок показника та квадрата величини зміщення оцінки, яке є різницею між очікуваною величиною оцінки (тобто середнім із вибірових оцінок) та дійсним значенням показника для генеральної сукупності.

Система статистичних показників – сукупність взаємопов'язаних статистичних показників, які всебічно й науково обґрунтовано відображають явища та процеси суспільно-економічного життя за певних умов місця й часу.

Стандартна похибка – середньоквадратичне відхилення вибірових оцінок показника за всіма можливими вибірками цього дизайну від дійсного значення показника для генеральної сукупності.

Статистична вага – кількість одиниць генеральної сукупності, які репрезентує (представляє) одна одиниця вибірки; величина, обернена до ймовірності відбору одиниці вибірки.

ПЕРЕДМОВА

В обстеженні умов життя домогосподарств (далі - ОУЖД) вимірюється комплекс соціально-економічних характеристик домогосподарств і населення України. Обстеження передбачає застосування різних типів статистичних формулярів для збирання інформації (запитальники основного та квартальних інтерв'ю, щоденники поточних витрат домогосподарств і анкети одноразових опитувань тощо), що вимагає реалізації складного поетапного процесу обробки результатів обстеження.

Методика розроблена з метою поліпшення поінформованості користувачів щодо технології отримання результатів ОУЖД.

Методика складається з трьох розділів і додатків. У першому розділі наведено загальну характеристику обстеження й огляд основних етапів розробки його результатів. У другому розділі наведено опис етапів обробки даних, що здійснюються на регіональному рівні (контроль заповнених первинних статистичних формулярів, уведення первинних даних у ПЕОМ, контроль і коригування введених даних обстеження). У третьому розділі розглянуто функції і завдання, що виконуються на центральному рівні Головним міжрегіональним управлінням статистики у м. Києві (ГМУС у м. Києві) (конвертація даних обстеження у формат SPSS, контроль, очищення даних, імпутація відсутніх даних обстеження, формування бази даних обстеження по окремих розділах інструментарію); Держкомстатом (формування файлів аналізу даних, формування системи статистичних ваг, аналіз якості, у т.ч. розрахунки характеристик надійності оцінок показників, та підготовка інформаційно-аналітичних матеріалів за результатами обстеження).

1. ЗАГАЛЬНІ ПОЛОЖЕННЯ

Вибіркове обстеження умов життя домогосподарств є основним джерелом інформації для комплексного дослідження рівня матеріального добробуту населення України. Це обстеження було запроваджено з 1999 року. Методологія обстеження в цілому відповідає загальноприйнятим міжнародним стандартам і вимогам до організації та проведення таких спостережень. Дані обстеження використовуються в системі національних рахунків, для розрахунку індексу споживчих цін, для моніторингу бідності, при розробці та моніторингу програм соціального захисту населення тощо.

З 2002 року показники споживчих витрат за результатами ОУЖД розраховуються відповідно до Міжнародної класифікації індивідуального споживання товарів і послуг за цілями (COICOP–HBS), рекомендованої Євростатом. Зазначена класифікація застосовується в обстеженнях витрат домогосподарств, а також для розрахунків макроекономічних показників щодо сектору домашніх господарств і при розрахунках індексу споживчих цін.

Починаючи з 2004 року, в органах державної статистики була впроваджена нова система організації проведення вибіркового обстеження населення (домогосподарств). Створена єдина уніфікована мережа фахівців з інтерв'ювання, які проводять одночасно три вибіркові обстеження населення (домогосподарств): умов життя домогосподарств, економічної активності населення та сільськогосподарської діяльності домогосподарств у сільській місцевості. ОУЖД здійснюється на окремій мережі респондентів – неінституційних домогосподарств. Щороку для участі в цьому обстеженні відбирається близько 13 тис. домогосподарств, які представляють усі регіони України, окрім I та II зон, що зазнали радіоактивного забруднення внаслідок Чорнобильської катастрофи. Формування вибірки здійснено на основі процедури стратифікованого багатоступеневого відбору. Вибірка територій діє протягом п'яти років, вибірка домогосподарств – одного року.

ОУЖД складається з трьох основних частин, для одержання інформації по кожній з яких застосовуються різні статистичні формуляри.

I. Збір даних щодо загальної характеристики домогосподарства – основне інтерв'ю. Опитування домогосподарства проводиться на початку обстеження за програмою відповідного запитальника стосовно загальних характеристик домогосподарства (форма № 2-УЖД):

- склад домогосподарства та його житлові умови (Розділ 1);
- наявність та використання земельних ділянок, худоби та птиці (Розділ 2);
- характеристики членів домогосподарства: антропометричні дані, рівні освіти, статуси зайнятості тощо (Розділ 3).

Крім того, під час опитування фахівець з інтерв'ювання заповнює контрольну картку складу домогосподарства (форма № 1-УЖД), у якій протягом усього періоду обстеження спостерігаються зміни в його складі.

II. Спостереження за витратами і доходами домогосподарств протягом кварталу. Здійснюється за допомогою двох інструментів обстеження:

- *Щоденник поточних витрат домогосподарств* (форма № 3-УЖД), у якому респонденти двічі на квартал протягом тижня (здійснюється розподіл респондентів на 6 ротацийних груп) щоденно записують усі витрати та надають їх детальний опис (наприклад, по куплених продуктах харчування – назва продукту, його вага та вартість, місце покупки). У щоденнику надається інформація щодо:

- купівлі продуктів харчування, алкогольних напоїв і тютюнових виробів (Розділ 1);
- купівлі продовольчих товарів і кормів для годування домашніх тварин (Розділ 2);
- витрат на харчування поза домом (Розділ 3);
- споживання продуктів, вироблених в особистому підсобному господарстві, подарованих або отриманих безкоштовно з інших джерел (Розділ 4);

– витрат на купівлю, виготовлення, ремонт непродовольчих товарів і на оплату послуг (Розділ 5).

• *Квартальний запитальник про витрати і доходи домогосподарств* (форма № 5-УЖД), за допомогою якого проводиться щоквартальне опитування домогосподарства у перший місяць після звітного кварталу. На цьому етапі здійснюється збір даних щодо значних і нерегулярних витрат, а також – щодо доходів домогосподарства:

- витрати на купівлю продуктів харчування для тривалого використання (Розділ 1);
- витрати на непродовольчі товари (Розділ 2);
- житлово-комунальні витрати (Розділ 3);
- земля для ведення особистого підсобного господарства; худоба птиця; самозаготівлі (Розділ 4);
- інші витрати (Розділ 5);
- індивідуальні доходи членів домогосподарства (Розділ 6) – збирається інформація окремо по кожному члену домогосподарства віком 16 років і старше;
- контроль повноти заповнення індивідуальних запитальників (Розділ 7);
- доходи домогосподарства (Розділ 8).

Оскільки домогосподарствам складно пригадати всі витрати та доходи за квартал, вони записують їх протягом кварталу в “Журнал квартальних витрат” (форма № 4-УЖД). Цей журнал використовується під час інтерв'ю в домогосподарстві тільки як допоміжний інструмент для заповнення основних анкет обстеження.

III. Одноразові тематичні опитування – під час квартального опитування за допомогою додаткових анкет проводяться тематичні опитування з питань:

• витрат домогосподарств на будівництво та ремонт житла і господарських будівель – Анкета № 1 (проводиться щокварталу разом із квартальним інтерв'ю);

• оцінки членами домогосподарств стану свого здоров'я та рівня доступності окремих видів медичної допомоги – 1 частина Анкети № 2 (проводиться щорічно в жовтні разом із квартальним інтерв'ю за III квартал);

• доступу домогосподарств до окремих товарів і послуг/ наявності товарів тривалого користування – 2 частина Анкети № 2 (проводяться в жовтні разом із квартальним інтерв'ю за III квартал, чергуючись між собою, тобто кожне з обстежень проводиться раз на 2 роки);

• самооцінки домогосподарствами рівня достатності своїх доходів і доступу до Інтернету – Анкетою № 3 (проводиться щорічно в січні разом із квартальним інтерв'ю за IV квартал).

Обробка результатів обстеження здійснюється відповідно до Технологічного плану державних статистичних спостережень на двох рівнях.

1). На регіональному (у головних управліннях статистики):

- візуальний контроль заповнених анкет, запитальників та інших статистичних формулярів – проводиться фахівцями з інтерв'ювання у домогосподарствах під час та після проведення опитування, а також спеціалістами-економістами регіональних підрозділів ОУЖД під час приймання анкет із результатами опитування від фахівців з інтерв'ювання;
- уведення даних обстеження в ПЕОМ – виконується засобами програмного забезпечення СУБД SQL Anywhere, для статистичних формулярів обстеження розроблені окремі комплекси електронної обробки даних (КЕОІ);
- контроль уведених у ПЕОМ даних обстеження – первинний, фрагментарний і глобальний контролю – реалізовано у програмному забезпеченні СУБД SQL Anywhere.

2). На центральному:

У ГМУС у м. Києві:

- завантаження даних регіональних фрагментів бази даних обстеження – паралельно проводиться контроль узгодженості даних регіональних фрагментів засобами програмного забезпечення СУБД SQL Anywhere;

- вивантаження загальної бази даних обстеження в середовище SPSS і проведення вторинного контролю засобами SPSS – логічний контроль між запитаннями та розділами інструментарію, контроль відповідності даних допустимим значенням;
- очищення даних – виправлення помилок, виявлених при проведенні контролів, агрегація даних по відповідних розділах щоденників і кварталних запитальників, вилучення дублюючих значень;
- імпутація відсутніх даних – створюються файли даних для пошуку "донорів" для імпутації методами "найближчого сусіда" та "середнього значення";
- формування файлів даних для передачі в Держкомстат – формуються стандартні набори файлів даних SPSS для кожного типу статистичного формуляра та передаються для подальшої обробки.

У Держкомстаті:

- логічний контроль і узгодження даних – логічний контроль між запитаннями та розділами статистичних формулярів, а також контроль узгодженості даних відповідних розділів різних формулярів;
- формування файлів аналізу даних – об'єднуються дані кварталного циклу обстеження, отримані за різними статистичними формулярами, формуються додаткові по відношенню до "прямих" змінних, наявних у статистичних формулярах, зокрема агреговані змінні доходів і витрат;
- формування системи статистичних ваг – ураховується загальна ймовірність відбору домогосподарства, реальні рівні участі певних типів домогосподарств у обстеженні, додаткова інформація стосовно певних демографічних характеристик домогосподарств, необхідність забезпечення єдиної системи ваг для осіб і домогосподарств;
- порівняння результатів обстеження в динаміці та із зовнішньою інформацією – перевірка узгодженості з відповідною інформацією системи національних рахунків, демографічної статистики тощо;
- оцінка надійності даних – розрахунок основних характеристик надійності даних вибіркового обстеження для агрегованих показників доходів і витрат;
- підготовка інформаційно-публікаційних матеріалів з результатами обстеження.

Послідовність основних етапів робіт з обробки даних ОУЖД наведена у додатку 1.

2. ОБРОБКА РЕЗУЛЬТАТІВ ОБСТЕЖЕННЯ НА РЕГІОНАЛЬНОМУ РІВНІ

2.1. Візуальний контроль заповнених первинних статистичних формулярів

Аналіз повноти, якості та логічної узгодженості заповнених первинних статистичних формулярів ОУЖД здійснюється візуально спеціалістами-економістами підрозділів ОУЖД головних управлінь статистики під час приймання від фахівців з інтерв'ювання анкет, щоденників і запитальників. Основна мета цієї роботи – виявлення помилок, серйозних недоліків у заповнених анкетах і запитальниках, а також усунення цих недоліків до початку робіт з уведення первинних даних у ПЕОМ.

Під час проведення контролю спеціаліст-економіст звертає увагу на:

- повноту, якість і логічну узгодженість заповнення формулярів;
- наявність однакових тенденцій у відповідях різних респондентів, опитаних одним і тим самим фахівцем з інтерв'ювання;
- наявність підтверджуючих підписів респондентів у випадку виправлення записів у анкетах і запитальниках;
- рівень участі домогосподарств в ОУЖД та причини відмов домогосподарств від участі в обстеженні;
- рівень збереження мережі домогосподарств в ОУЖД.

На основі цього аналізу виявляються помилки та з'ясовується їх характер. Виявлені помилки за характером можуть бути випадковими (унаслідок неувважності фахівця з інтерв'ювання) або систематичними (унаслідок неправильного розуміння методологічних

вимог до проведення ОУЖД). Контролер звертає особливу увагу на систематичні помилки, що носять повторний характер (це помилки, допущені фахівцем з інтерв'ювання у попередніх періодах, проте він продовжує їх повторювати).

Аналіз наявності однакових тенденцій у відповідях різних респондентів, опитаних одним і тим самим фахівцем з інтерв'ювання, та логічної обґрунтованості цих тенденцій здійснюється з метою виявлення випадків фальсифікації даних фахівцем з інтерв'ювання або його впливу на респондентів під час опитування. Такий аналіз проводиться шляхом візуального порівняння відповідної інформації, отриманої фахівцем з інтерв'ювання у звітному періоді під час опитування декількох домогосподарств.

Контролер визначає по кожному фахівцю з інтерв'ювання рівень участі домогосподарств в ОУЖД та рівень збереження мережі домогосподарств, для визначення якого розраховується кількість домогосподарств, які вибули з обстеження у звітному періоді (без урахування нежилих квартир, нежилих приміщень), а також узагальнює та здійснює аналіз причин відмов респондентів від участі в обстеженні.

Контролер фіксує в анкеті або запитальнику виявлені помилки або запитання, що виникли під час перевірки. Відповідно до характеру помилки дані уточнюються у фахівця з інтерв'ювання або безпосередньо у респондентів. Усі виправлення в анкетах, щоденниках і запитальниках мають бути обов'язково підтверджені підписами респондентів.

2.2. Уведення первинних даних обстеження до ПЕОМ

Уведення первинних даних до ПЕОМ здійснюють спеціалісти-економісти підрозділів ОУЖД головних управлінь статистики з використанням програмного забезпечення, створеного для архітектури "клієнт-сервер" засобами програмного забезпечення Sybase-Powersoft на основі СУБД SQL Anywhere.

Програмне забезпечення розроблено на засадах системного підходу до організації проведення обстеження з одночасним взаємозв'язком з різними статистичними формулярами ОУЖД. Застосовуються такі КЕОІ:

- 1). Графік роботи фахівця з інтерв'ювання.
- 2). Контрольна картка складу домогосподарства.
- 3). Запитальник основного інтерв'ю.
- 4). Щоденник поточних витрат і Квартальний запитальник про витрати і доходи домогосподарства.

База даних (далі – БД) розміщується на сервері, що забезпечує можливість доступу з декількох клієнтських станцій. З центрального рівня на регіональний на початку проведення нового циклу обстеження направляється класифікатор КОАТУУ, визначається перелік територіальних одиниць регіону, які увійшли до вибірки (термін дії територіальної вибірки становить *п'ять* років). На початку *четвертого* кварталу року, що передує звітному, формується каталог ПТОВ, на основі даних територіальної вибірки. На центральному рівні у вересні місяці року, що передує обстежуваному, визначається дата початку проведення основного та квартального інтерв'ю.

На початку проведення обстеження заповнюється каталог фахівців з інтерв'ювання, який при появі змін у списку фахівців з інтерв'ювання поновлюється протягом усього періоду обстеження. Протягом дії територіальної вибірки коди фахівців з інтерв'ювання не змінюються в рамках ПТОВ. Каталог фахівців з інтерв'ювання не може поповнюватись у процесі обстеження. На центральний рівень передається тільки один раз разом із основним інтерв'ю.

У *четвертому* кварталі року, що передує звітному, визначається список домогосподарств, які будуть обстежуватись. Використовуючи дані вибіркової сукупності домогосподарств на регіональному рівні, заповнюється каталог домогосподарств. Коди домогосподарств не підлягають зміні або заміні в процесі обробки. На центральний рівень каталог передається разом із основним інтерв'ю. Після передачі даних на центральний рівень каталог домогосподарств не може поповнюватись. Також на початок обстежуваного року проводиться розподіл домогосподарств на ротаційні групи для ведення щоденників поточних витрат і формування графіка відвідування фахівцем з інтерв'ювання домогосподарств для проведення основного та квартальних інтерв'ю.

Графік роботи фахівця з інтерв'ювання

Програмне забезпечення, розроблене для складання графіка роботи фахівця з інтерв'ювання, є першим елементом комплексної системи електронної обробки результатів ОУЖД. За допомогою цього програмного забезпечення спеціалісти-економісти головних управлінь статистики виконують такі функції:

- Уведення та коригування даних каталогів (каталог фахівців з інтерв'ювання, каталог ПТОВ, каталог домогосподарств, графік відвідування домогосподарств для проведення основного та квартальних інтерв'ю тощо).
- Перегляд даних каталогів.
- Формування запитів на пошук необхідної інформації в каталогах.
- Формування запитів на вивантаження-завантаження даних.
- Друк графіків роботи фахівця з інтерв'ювання.

Функції, що виконуються на сервері:

- Автоматичне кодування адресної частини домогосподарств.
- Логічний контроль каталогів.
- Вилучення фрагментів класифікатора КОАТУУ інших регіонів.
- Вилучення фрагментів каталогу ПТОВ інших регіонів.
- Забезпечення цілісності та узгодженості каталогів і довідників.

Інформація розробляється щоквартально. Кількість кварталів повного циклу обробки – 5. Обробка інформації виконується щоквартально в поточному році та один раз у період проведення основного інтерв'ю в році, що передує звітному. На регіональному рівні складання графіків роботи фахівців з інтерв'ювання повинно виконуватись протягом двох тижнів з 1 по 15 число останнього місяця кварталу (десять робочих днів).

Контрольна картка складу домогосподарства

Програмне забезпечення, розроблене для введення в ПЕОМ первинних даних контрольної картки складу домогосподарств, є другим елементом комплексної системи електронної обробки результатів ОУЖД. Основне опрацювання даних виконується на сервері, на робочій станції підтримується графічний інтерфейс користувача для доступу до БД.

Функції клієнтської частини:

- введення та коригування первинних даних контрольної картки домогосподарства по членах домогосподарства;
- перегляд даних каталогу домогосподарств;
- модифікація каталогу членів домогосподарств;
- перегляд та друк протоколу контролю контрольної картки домогосподарства;
- формування запитів на проведення контролю;
- формування запитів на формування довідок про стан БД;
- друк вихідних документів (довідки та протоколи).

Функції, які виконуються на сервері:

- контроль узгодженості складу домогосподарства відповідного каталогу;
- контроль по домогосподарству відповідного каталогу з показниками графіка роботи фахівця з інтерв'ювання;
- глобальний контроль, формування протоколу;
- забезпечення цілісності й узгодженості каталогів і довідників.

Періодичність обробки матеріалів – річна, квартальна (кількість кварталів повного циклу обробки – 5). Обробка інформації виконується щоквартально в поточному році та один раз у період проведення основного інтерв'ю в році, що передує звітному. Тривалість обробки – 40 днів.

Порядок збору інформації та схема її проходження:

1) контрольні картки заповнюються в домогосподарстві під час проведення основного інтерв'ю і актуалізуються під час квартальних інтерв'ю;

2) заповнені бланки контрольних карток разом із запитальниками доставляються в регіональні підрозділи ОУЖД;

3) дані контрольних карток вводяться до БД, де накопичуються, коригуються, контролюються;

4) на основі даних контрольних карток домогосподарств формується каталог членів домогосподарств;

5) контроль даних контрольних карток домогосподарств виконується за запитом користувача, глобальний контроль виконується в рамках глобального контролю запитальників основного та квартального інтерв'ю;

6) дані контрольних карток на державний рівень передаються разом із даними запитальників основного та квартального інтерв'ю.

Запитальник основного інтерв'ю

Програмне забезпечення, розроблене для введення в ПЕОМ первинних даних основного інтерв'ю домогосподарства, є третім елементом комплексної системи електронної обробки результатів ОУЖД. Основне опрацювання даних виконується на сервері, на робочій станції підтримується графічний інтерфейс користувача для доступу до БД.

Функції клієнтської частини:

- уведення та коригування даних запитальника основного інтерв'ю по домогосподарству та членах домогосподарства;
- перегляд даних каталогу домогосподарств і членів домогосподарств;
- перегляд і друк протоколу контролю;
- формування запитів на проведення контролю та вивантаження;
- формування запитів на формування довідок про стан БД;
- друк вихідних документів (довідки та протоколи).

Функції, які виконуються на сервері:

- контроль описів основного інтерв'ю по члену домогосподарства;
- контроль описів основного інтерв'ю по домогосподарству;
- глобальний контроль, формування протоколу;
- забезпечення цілісності та узгодженості каталогів, таблиць із показниками основного інтерв'ю та довідників;
- формування фрагментів для відсилки на державний рівень;
- формування довідок про стан БД;
- формування файлів txt- формату для вивантаження в інші середовища;
- коригування даних основного інтерв'ю та каталогу домогосподарств за результатами основного інтерв'ю, формування протоколу;
- вивантаження каталогів.

Періодичність обробки матеріалів – річна. Тривалість обробки – 40 днів.

Порядок збору інформації та схема її проходження:

1) запитальники основного інтерв'ю подаються до регіональних підрозділів ОУЖД фахівцями з інтерв'ювання в міру завершення опитування домогосподарств;

2) дані основного інтерв'ю вводяться до БД, де накопичуються, контролюються та коригуються, у т.ч. в разі необхідності вручну (у частині окремих показників);

3) після закінчення введення до ПЕОМ усіх запитальників проводиться перевірка повноти надходження даних обстеження та повторний (після коригування) контроль;

4) формуються довідки про стан БД та протоколи контролів;

5) дані основного інтерв'ю вивантажуються із БД до файлів текстового формату у вигляді регіональних фрагментів БД і передаються засобами електронної пошти на державний рівень у ГМУС у м.Києві.

Щоденник поточних витрат домогосподарства та Квартальний запитальник про витрати і доходи домогосподарства

Програмне забезпечення, розроблене для введення в ПЕОМ первинних даних щоденників поточних витрат і квартального запитальника про витрати і доходи

домогосподарства, є останнім елементом комплексної системи електронної обробки результатів ОУЖД. Основне опрацювання даних виконується на сервері, на робочій станції підтримується графічний інтерфейс користувача для доступу до БД.

Функції клієнтської частини (клієнтська станція):

- уведення, коригування та перегляд даних щоденників поточних витрат домогосподарств;
- уведення, часткове коригування та перегляд даних квартальних запитальників про витрати і доходи домогосподарств;
- формування запитів на коригування каталогів обстеження по квартальних даних обстеження;
- візуалізація стану обробки каталогів за наслідками контролю;
- перегляд і друк класифікатора товарів і послуг, довідника граничних значень, протоколу контролю.

Функції, які виконуються на сервері:

- фрагментарний контроль щоденників поточних витрат домогосподарств і квартального запитальника про витрати і доходи домогосподарств, формування протоколів контролю;
- глобальний контроль щоденників поточних витрат і запитальника квартального інтерв'ю, формування протоколів контролю;
- суцільний і фрагментарний контроль кодифікатора товарів і послуг і довідника граничних значень;
- забезпечення цілісності й узгодженості каталогів і довідників;
- коригування даних запитальників квартального інтерв'ю;
- контроль описів квартального інтерв'ю по члену домогосподарства;
- контроль описів щоденників поточних витрат квартального інтерв'ю по домогосподарству;
- формування вивантаженого фрагмента щоденників поточних витрат домогосподарств і фрагмента квартальних запитальників про витрати і доходи домогосподарств і змін по контрольних картках складу домогосподарств за поточний квартал для відправлення на державний рівень;
- вивантаження та завантаження коректур кодифікатора товарів і послуг і довідника граничних значень (надходять централізовано);
- вивантаження БД для передачі.

Періодичність обробки матеріалів – квартальна. Повний цикл обробки містить 4 квартали. Тривалість обробки – 40 днів.

Порядок збору інформації та схема її проходження:

1) щоденники поточних витрат та запитальники квартального інтерв'ю подаються до регіональних підрозділів ОУЖД фахівцями з інтерв'ювання в міру завершення опитування домогосподарств. Одним домогосподарством за обстежуваний квартал заповнюється два щоденники поточних витрат (за два тижні) й один запитальник квартального інтерв'ю;

2) обробка квартальної інформації виконується у два етапи:

на першому – обробка даних щоденників поточних витрат;

на другому – обробка змін до контрольних карток і обробка запитальників квартального інтерв'ю.

Для кожного з етапів виконуються такі процедури:

- Дані щоденників / запитальників квартального інтерв'ю вводяться в ПЕОМ до БД, де накопичуються, контролюються та коригуються, у т.ч. в разі необхідності, вручну (у частині окремих показників).

- Після введення до ПЕОМ здійснюється перевірка повноти надходження даних обстеження та повторний (після коригування) контроль.

- Формуються протоколи контролю та довідки про стан БД.

- Дані щоденників / запитальників квартального інтерв'ю вивантажуються із БД до файлів текстового формату у вигляді регіональних фрагментів БД для подальшої передачі на державний рівень у ГМУС у м. Києві засобами електронної пошти.
- Разом із даними квартального інтерв'ю вивантажуються зміни до контрольних карток домогосподарств (інформація по нових членах окремо від змін у складі домогосподарства) та регіональна довідка про зміни в каталозі членів домогосподарств протягом обстеження.

Одноразові опитування домогосподарств

Обробка даних одноразових опитувань проводиться за схемою обробки даних квартального інтерв'ю.

Опитування щодо витрат на будівництво та ремонт – *щоквартально* разом із даними квартального інтерв'ю (*4 квартали*).

Опитування щодо стану здоров'я членів домогосподарства – *один раз на рік* разом із даними квартального інтерв'ю за *третій* квартал.

Опитування щодо наявності у домогосподарствах товарів тривалого користування – *один раз на два роки* разом із даними квартального інтерв'ю за *третій* квартал.

Опитування щодо доступу домогосподарств до окремих видів товарів та послуг – *один раз на два роки* разом із даними квартального інтерв'ю за *третій* квартал.

Опитування щодо самооцінки домогосподарствами рівня своїх доходів та доступу до Інтернету – *один раз на рік* разом із даними квартального інтерв'ю за *четвертий* квартал.

Порядок збору інформації та схема її проходження:

1) анкети опитування фахівці з інтерв'ювання подають до регіональних підрозділів ОУЖД разом із запитальниками квартального інтерв'ю в міру завершення квартального опитування домогосподарств;

2) дані анкет вводяться до БД, де накопичуються, контролюються та коригуються, у т.ч. разі необхідності вручну (у частині окремих показників);

3) проводиться перевірка повноти надходження даних обстеження та повторний (після коригування) контроль;

4) формуються протоколи контролю та довідки про стан БД;

5) дані одноразових опитувань разом із даними відповідного квартального інтерв'ю вивантажуються із БД до файлів текстового формату у вигляді регіональних фрагментів БД для подальшої передачі на державний рівень у ГМУС у м. Києві засобами електронної пошти.

2.3. Контроль первинних даних обстеження

При обробці статистичних формулярів ОУЖД передбачається здійснення контролю за такими етапами:

1) *Первинний контроль* – виконується безпосередньо під час введення даних до ПЕОМ і викликає або блокування помилкових даних (без виправлення помилки неможливе подальше введення даних запису) або вимагає внесення виправлення (після підтвердження первинними даними). Конкретний варіант дій залежить від запитання й типу помилки.

2) *Фрагментарний контроль* – виконується після закінчення етапу введення або коригування даних розділу інструментарію. Проводиться перевірка лише конкретного розділу без урахування стану інших розділів інструментарію.

3) *Глобальний контроль* – виконується для кожного інструментарію в цілому та складається з:

- усіх груп перевірок фрагментарного контролю,
- перехресного контролю розділів інструментарію,
- контролю повноти заповнення документа,
- перевірок узгодженості документа з іншими видами інструментарію.

Цей вид контролю виконується як для перевірки окремого домогосподарства, так і групи домогосподарств за діапазоном кодів домогосподарств, що дозволяє виконувати контроль у цілому по регіону.

Тільки після проведення глобального контролю для кожного статистичного формуляра приймається рішення щодо формування регіональних фрагментів БД.

При обробці даних обстеження визначаються помилки двох типів:

– помилки жорсткого контролю – помилки, які потребують обов'язкового виправлення;

– помилки попереджувального контролю – помилки у введених до ПЕОМ даних та ситуації, що не являються помилкою у цьому випадку. Після порівняння введених даних із первинними даними, наведеними у статистичних формулярах, проводиться коригування введених даних лише у випадку наявності помилки введення.

Нижче наведені описи основних видів контролю, що здійснюються в рамках окремих КЕОІ.

Графік роботи фахівця з інтерв'ювання

Під час заповнення каталогу фахівців з інтерв'ювання та каталогу домогосподарств інформація контролюється лише в інтерактивному режимі під час її запису до бази даних:

- відповідність коду ПТОВ значенню в каталозі ПТОВ;
- відповідність коду фахівця з інтерв'ювання значенню в каталозі фахівців з інтерв'ювання;
- відповідність коду вулиці значенню в довіднику вулиць;
- якщо по домогосподарству надані основні дані, то по цьому ж домогосподарству повинні бути й додаткові відомості (крім прізвища);
- якщо по домогосподарству немає основних даних, то не має бути й додаткових відомостей.

Каталог фахівців з інтерв'ювання коригується на регіональному рівні при появі змін у списку фахівців з інтерв'ювання протягом усього періоду обстеження. У каталозі по фахівцю з інтерв'ювання передбачається коригування всіх даних, крім коду.

Контрольна картка складу домогосподарства

Для контрольної картки складу домогосподарства передбачається здійснення контролю по таких етапах:

- первинний контроль при клавіатурному введенні даних у ПЕОМ у регіональних підрозділах ОУЖД;
- глобальний перехресний логічний контроль уведених даних контрольної картки складу домогосподарства в межах домогосподарства;
- по кожному домогосподарству глобальний логічний контроль відповідності даних каталогу членів домогосподарств іншим основним вхідним формам інструментарію.

Первинний контроль інформації складається з перевірок:

- відповідності шаблону даних,
- входження значення до встановленого діапазону мінімальних і максимальних значень.

З метою недопущення введення кодових значень, які не відповідають складу відповідного довідника, користувачу надається можливість вибрати одне значення з відповідного довідника в інтерактивному режимі. При встановленні невідповідності значень кодів родинних стосунків та сімейного стану виконується блокування подальшої обробки контрольної картки.

Глобальний перехресний логічний контроль полягає у проведенні перевірки:

- логіки заповнення показників по кожному члену домогосподарства окремо;
- узгодженості складу домогосподарства;
- по кожному члену домогосподарства узгодженості віку з іншими показниками для цього члена домогосподарства;
- узгодженості причин зміни у складі домогосподарства з результатами обробки даних по контрольних картках на попередніх етапах.

Після введення інформації по кожному домогосподарству обов'язково проводиться глобальний контроль на регіональному рівні відповідності даних каталогу членів домогосподарств іншим основним вхідним формам інструментарію.

Порушення узгодженості каталогів домогосподарств і членів домогосподарств блокується на етапі введення контрольної картки шляхом недопущення введення контрольних карток до відсутніх домогосподарств.

При виявленні помилок при введенні даних контрольної картки складу домогосподарства складається протокол контролю. Згідно з протоколом уносяться виправлення в первинні звіти і знову проводяться всі види контролю.

У разі необхідності виправлення помилок на етапі обробки основного інтерв'ю коригування контрольної картки передбачає відповідні зміни всіх показників контрольної картки, окрім кодів змін у складі домогосподарства, уведення нових членів домогосподарства.

Запитальник основного інтерв'ю

При введенні в ПЕОМ запитальника основного інтерв'ю обов'язково виконується первинний контроль:

- перевірка відповідності шаблону даних (структурі статистичного формуляра);
- перевірка відповідності основних відомостей графіка роботи фахівця з інтерв'ювання (тільки для титульного листа запитальника основного інтерв'ю);
- перевірка входження значення відповіді до встановленого діапазону;
- перевірка узгодженості відповіді з попереднім та наступним значенням (тільки розділи 1-3 запитальника);
- перевірка узгодженості відповіді з даними Контрольної картки (тільки розділ 3 запитальника).

При введенні даних розділів 1-3 запитальника помилкові значення блокуються. Для виправлення помилок узгодженості поточної відповіді з попередніми надається можливість повернутися на потрібне запитання (всі раніше введені до цього питання відповіді скасовуються).

Дублюючі записи при введенні блокуються засобами контролю БД – не допускається наявності в БД двох записів з однаковими ключовими показниками.

Уведення даних основного інтерв'ю в частині даних по членах домогосподарств дозволяється лише для тих домогосподарств, для яких заповнена контрольна картка.

Після введення та коригування даних обов'язково проводиться глобальний контроль, під час якого виконуються:

- контроль каталогу домогосподарств – виконується блок перевірок для оброблення графіка роботи фахівця з інтерв'ювання;
- контроль каталогу членів домогосподарств – виконується блок перевірок для оброблення контрольної картки щодо складу домогосподарства, наявності дітей тощо;
- контроль основних відомостей титульного листа запитальника основного інтерв'ю;
- контроль узгодженості відповідей у розділах 1-3 запитальника основного інтерв'ю;
- контроль узгодженості відповідей у розділах 1-3 запитальника основного інтерв'ю із графіком роботи фахівця з інтерв'ювання, контрольною картою, основними відомостями титульного листа запитальника основного інтерв'ю;
- контроль повноти заповнення запитальника основного інтерв'ю.

За результатами глобального контролю складається протокол контролю основного інтерв'ю та контрольної картки. Дані в протоколі глобального контролю подаються по зростанню:

- 1) номерів домогосподарств;
- 2) номерів членів домогосподарства;
- 3) номеру повідомлення про помилку.

За результатами глобального контролю домогосподарства та члени домогосподарств позначаються залежно від типу виявленої помилки (зірочка – помилка жорсткого контролю, рисочка – помилка попереджувального контролю). Згідно з протоколом уносяться необхідні виправлення відповідно до встановленого порядку і знову проводяться всі види контролю.

Щоденник поточних витрат домогосподарства та Квартальний запитальник про витрати і доходи домогосподарства

При обробці даних щоденників поточних витрат і запитальників квартального інтерв'ю на регіональному рівні передбачається здійснення контролю в три етапи:

- 1). Первинний контроль при безпосередньому введенні даних у ПЕОМ.
- 2). Фрагментарний контроль уведеного розділу, у т.ч. титульного листа запитальника квартального інтерв'ю, груп «інших розділів» і одноразових анкет по домогосподарству та індивідуальних запитальників по кожному дорослому члену домогосподарства.
- 3). Глобальний контроль по домогосподарству або групі домогосподарств – перехресний логічний контроль різних розділів і перевірка узгодженості з іншими формами інструментарію.

Арифметичний і балансовий контроль не передбачаються.

Під час *первинного контролю* проводиться:

- перевірка відповідності шаблону даних (структурі статистичного формуляра);
- перевірка відповідності основних відомостей графіку роботи фахівця з інтерв'ювання (тільки для щоденників поточних витрат);
- перевірка входження значення відповіді до встановленого діапазону;
- перевірка узгодженості відповіді з попереднім і наступним запитанням (тільки для розділів запитальника квартального інтерв'ю, починаючи з третього, й анкет);
- перевірка узгодженості відповіді з даними контрольної картки (тільки для розділів запитальника квартального інтерв'ю, починаючи з третього, й анкет).

Запитальник квартального інтерв'ю для членів домогосподарств уводиться лише при наявності заповненої контрольної картки.

Для кожного товару або послуги, тобто кожного рядка кожної таблиці щоденника поточних витрат домогосподарства, проводиться блок перевірок по показниках кількості, вартості, ціни у частині входження до встановлених діапазонів, а також щодо можливих місць придбання товарів і отримання послуг.

Товари або послуги, значення показників яких викликали повідомлення контролю, змінюють колір у рядках екранної форми. Підсвічення полів відбувається згідно з таблицею відповідності кольорів і позначень помилок.

При введенні даних щоденників поточних витрат і запитальників квартального інтерв'ю виконується жорсткий і попереджувальний контроль. В обох випадках виникнення помилки введення супроводжується зміною кольору на помилкових показниках сторінок щоденника поточних витрат, титульного листа й розділів 1 і 2 запитальника квартального інтерв'ю.

Під час *фрагментарного контролю* виконуються:

- перевірка відповідності даних титульного листа запитальника квартального інтерв'ю графіку роботи фахівця з інтерв'ювання,
- контроль узгодженості показників кількості та вартості з довідником граничних значень,
- контроль узгодженості відповідей у межах одного розділу (для титульного листа, розділів, починаючи з третього, усіх видів анкет запитальника квартального інтерв'ю),
- частковий контроль узгодженості даних титульного листа запитальника квартального інтерв'ю з даними каталогу домогосподарств.

Контроль даних щоденників поточних витрат

Щоденники поточних витрат уводяться тільки по тих домогосподарствах, що були занесені до каталогу при складанні графіку роботи фахівця з інтерв'ювання для основного інтерв'ю та по яких на етапі проведення основного інтерв'ю або попереднього квартального інтерв'ю не була встановлена ознака відмови від обстеження. Домогосподарства, які відмовилися від обстеження, виключаються з дерева перегляду і тим самим для них блокується введення щоденників поточних витрат.

Для заповнених щоденників виконується первинний контроль титульного листа щоденника поточних витрат домогосподарства, первинний контроль по днях заповнення щоденника, а також виконується група перевірок для кожного розділу щоденника.

Кожний показник, що вводиться в таблиці щоденника, перевіряється:

- 1) на наявність у довіднику граничних значень товарів і послуг разом із установленою одиницею виміру;

2) якщо в довіднику граничних значень товарів і послуг установлені межі кількості або вартості (мінімальне та максимальне значення), то введена кількість або вартість товару / послуги перевіряється на входження в цей діапазон;

3) у разі перевищення верхньої межі границі кількості або вартості додатково проводиться перевірка на відповідність коефіцієнту перевищення граничного значення кількості (або відповідно вартості).

Попередження про можливу помилку видаються у вигляді протоколу.

Глобальний контроль запитальника квартального інтерв'ю складається із сукупності перевірок запитальника квартального інтерв'ю та змін до контрольної картки домогосподарства:

- контроль каталогу домогосподарств,
- контроль каталогу членів домогосподарства,
- усі блоки фрагментарного контролю на титульний лист і розділи запитальника квартального інтерв'ю,
- контроль повноти й узгодженості заповнення запитальника квартального інтерв'ю,
- контроль взаємозв'язків показників запитальника квартального інтерв'ю з каталогами домогосподарств і членів домогосподарств.

Одноразові опитування домогосподарств

Для одноразових анкет перевірки їх заповнення виконується залежно від типу анкети, наявності її формального опису для звітнього кварталу, віку особи, який розраховується за даними каталогу членів домогосподарств, та коду зміни в складі домогосподарства.

Усі виявлені помилки відображаються у протоколі глобального контролю щоденників поточних витрат та запитальника квартального інтерв'ю. Дані в протоколі глобального контролю подаються за зростанням:

- 1) номерів домогосподарств;
- 2) номерів членів домогосподарства;
- 3) номеру повідомлення про помилку;
- 4) кодів товарів / послуг і одиниць виміру (для товарів / послуг).

За результатами глобального контролю домогосподарства та члени домогосподарств позначаються залежно від типу виявленої помилки (зірочка – помилка жорсткого контролю, рисочка – помилка попереджувального контролю). Згідно з протоколом уносяться необхідні виправлення відповідно до встановленого порядку і знову проводяться всі види контролю.

3. ОБРОБКА РЕЗУЛЬТАТІВ ОБСТЕЖЕННЯ НА ЦЕНТРАЛЬНОМУ РІВНІ

3.1. Функції та завдання, що виконуються в Головному міжрегіональному управлінні статистики у м. Києві

У ГМУС у м. Києві виконуються такі етапи обробки даних ОУЖД: формування єдиної бази даних обстеження, контроль, вивантаження бази даних у формат SPSS, логічний контроль, чищення та імпутація даних. На цьому рівні проводиться обробка даних усіх статистичних формулярів ОУЖД. Усі роботи виконуються в терміни, визначені Технологічним планом державних статистичних спостережень (див. додаток 2):

- дані контрольної картки надходять і обробляються разом із даними запитальника основного інтерв'ю, зміни до контрольних карток надходять разом із даними запитальників квартального інтерв'ю;
- дані опитування домогосподарств про витрати на будівництво та ремонт житла та господарських будівель (Анкета №1) надходять і обробляються *щоквартально* разом із даними запитальників кожного квартального інтерв'ю;
- дані опитування щодо оцінки членами домогосподарств стану свого здоров'я та рівня доступності окремих видів медичної допомоги (Анкета №2) надходять і обробляються разом із даними запитальника квартального інтерв'ю за *третій* квартал;
- дані опитувань щодо доступу домогосподарств до окремих товарів і послуг і наявності товарів тривалого користування (Анкета №2), які чергуються між собою,

- оскільки проводяться один раз на два роки, надходять і обробляються разом із даними запитальника квартального інтерв'ю за *третьої* квартал;
- дані опитування щодо самооцінки домогосподарствами рівня своїх доходів і доступу до Інтернету (Анкета №3) надходять і обробляються разом із даними запитальника квартального інтерв'ю за четвертий квартал.

3.1.1. Формування єдиної бази даних обстеження на основі регіональних фрагментів

Відповідно до Технологічного плану державних статистичних спостережень до ГМУС у м. Києві з головних управлінь статистики надходять у текстовому форматі масиви даних із результатами обстеження. Після цього відбувається формування єдиної бази даних ОУЖД, а саме: запис первинних даних до бази даних, контроль і вивантаження з СУБД SQL Anywhere в статистичний пакет SPSS інформації по кожному статистичному формуляру.

Контроль при завантаженні регіональних фрагментів

Регіональні фрагменти бази, які надходять у текстовому форматі, завантажуються до загальної БД центрального рівня, де накопичуються та контролюються. Під час формування загальної БД обстеження завантаження регіональних фрагментів бази може блокуватися при наявності неузгодженостей даних регіональних фрагментів або невідповідностей у каталогах домогосподарств і членів домогосподарств на регіональному та державному рівнях. У разі блокування проводиться контроль на виявлення причини, відновлюється стан БД до завантаження та відправляється запит до відповідного регіонального підрозділу на повторну відправку відкоригованого регіонального фрагмента.

Можливі такі випадки порушення узгодження даних регіонального фрагмента:

- Наявність кодів фахівця з інтерв'ювання в регіональному каталозі домогосподарств, які відсутні як у каталогах фахівців з інтерв'ювання державного рівня, так і у відповідних регіональних фрагментах.
- Наявність даних по домогосподарствах регіонального фрагмента, які відсутні в регіональних каталогах домогосподарств.
- Наявність у регіональному каталозі членів домогосподарств даних по домогосподарствах, які відсутні в регіональному каталозі домогосподарств.
- Наявність (відсутність) даних по членах домогосподарства регіонального фрагмента, які відсутні (наявні) в регіональних каталогах членів домогосподарств.

Головна мета контролю при завантаженні регіональних фрагментів – визначення причин блокування завантаження внаслідок неузгодженості даних фрагментів або розбіжностей у каталогах домогосподарств і членів домогосподарства на регіональному та державному рівнях із подальшим виправленням неузгодженостей і забезпеченням належної повноти та якості інформації.

Порядок (схема) проходження інформації:

- 1). Регіональні фрагменти завантажуються до загальної БД державного рівня, де накопичуються та контролюються.
- 2). При виникненні неузгодженостей даних регіональних фрагментів завантаження блокується системою.
- 3). Відправляються запити на регіональний рівень для виправлення неузгодженостей.
- 4). Після виправлень регіональні фрагменти надсилаються на ГМУС у м. Києві, де вони завантажуються до загальної БД.
- 5). Дані вивантажуються з БД до файлів текстового формату для подальшої конвертації в програмне середовище SPSS.
- 6). Загальна БД зберігається в СУБД SQL Anywhere в первинному вигляді, а коригування, чищення та імпутація даних обстеження проводиться в середовищі SPSS.

3.1.2. Конвертація даних обстеження у формат SPSS, їх контроль і очищення

Після проведення контролів на узгодження регіональних фрагментів дані вивантажуються в середовище програмного забезпечення SPSS. Подальша обробка даних відбувається у цьому статистичному пакеті з використанням програми EXCEL для окремих операцій.

У середовищі SPSS також проводиться етап контролю даних, що зумовлене потужними можливостями цього пакета щодо побудови логічних залежностей і необхідністю додаткової перевірки помилок обробки даних, які могли виникнути або не були виявлені на попередніх етапах:

1). Логічний контроль заповнення статистичних формулярів – проводиться з використанням синтаксис-програм, розроблених особисто кожним спеціалістом-економістом, який займається обробкою конкретних формулярів чи їх частин. Перевіряється правильність заповнення певного розділу відповідно до логічних переходів і вказівок щодо заповнення формулярів.

2). Проводиться контроль даних на допустимість значень:

- відповідність одиниць виміру кодам товарів;
- відповідність і логічна узгодженість кодів товарів / послуг і кодів місць їх придбання;
- порівняння обсягів спожитого товару з нормами споживання;
- вартість товару / послуги порівнюється з даними довідника граничних значень, який формується на основі реальних даних за попередні періоди обстеження;
- відповідність кількості та вартості товарів / послуг.

3). Контроль на узгодженість і логічний контроль між даними різних розділів статистичних формулярів.

4). Контроль на узгодженість і логічний контроль між даними відповідних розділів щоденника поточних витрат і квартального запитальника:

- *перший* розділ щоденника та *перший* розділ квартального запитальника;
- *п'ятий* розділ щоденника та *другий, третій, четвертий* і *п'ятий* розділи квартального запитальника.

5). Контроль на узгодженість і логічний контроль між даними відповідних розділів запитальника основного інтерв'ю та квартального запитальника:

- *другий* розділ запитальника основного інтерв'ю та *четвертий* розділ квартального запитальника;
- *перший* розділ запитальника основного інтерв'ю та *другий* і *третій* розділи квартального запитальника;

6). Відбувається перевірка відповідності кодів товарів і послуг і записів у полях коментарів фахівців з інтерв'ювання.

За допомогою синтаксис-програми проводиться аналіз даних розділу статистичних формулярів та при виявленні помилок формується запит, який відправляється в регіональний підрозділ ОУЖД. На основі уточненої інформації, що надходить із регіонального рівня, проводиться очищення даних і виправлення помилок. У запитах для уточнення виявлених помилкових даних подається інформація про тип виявленої помилки. Помилкові дані записуються у порядку зростання:

- номерів домогосподарств;
- номерів членів домогосподарства;
- кодів товарів / послуг та одиниць виміру (для товарів / послуг);
- номеру повідомлення про помилку.

Процес контролю й очищення даних усіх типів статистичних формулярів (запитальника основного інтерв'ю, кожного розділу квартального запитальника та щоденника поточних витрат, анкет одноразових опитувань) відбувається за загальною схемою, наведеною в додатку 3.

Обробка даних на цьому етапі виконується в напівавтоматизованому режимі з використанням синтаксис-файлів SPSS, які можуть запускатися на виконання цілком або частинами по окремих змінних. Спеціалісти-економісти аналізують повідомлення програм про можливі помилки, надсилають запити у регіональні підрозділи для уточнення інформації та виправляють помилки на основі даних, отриманих від регіонів, які уточнюють інформацію в респондентів.

На кожному етапі контролю й очищення даних виводиться інформація про помилки, яку аналізує спеціаліст-економіст, уточнюється та за потреби коригується. Цей процес є *ітераційним* і повторюється до тих пір, поки не будуть усунені всі виявлені помилки.

Процес очищення даних обстеження можна умовно розділити на етапи:

Перший – підготовчі роботи. Створюються робочі файли даних, тобто створюються нові змінні та додаються дані з довідників; виявлені помилки кодуються за типами; формуються файли коментарів. За результатами аналізу файла коментарів і уточнення некоректної інформації спеціаліст-економіст виконує відповідні виправлення.

Другий – перевірка значень змінних усіх файлів на допустимість. Значення кожної змінної порівнюється з відповідними довідниками. Інформацію з виявленими невідповідностями аналізує спеціаліст-економіст та відправляє для уточнення на регіональний рівень. Після отримання результатів уточнення даних спеціаліст-економіст виправляє всі невідповідності в даних і проводить аналіз повторно.

Третій – виконується аналіз кодів товарів та виявляються дублюючі записи у відповідних розділах щоденників поточних витрат домогосподарств і квартальних запитальників. Спеціаліст-економіст аналізує результати роботи програми, вносить необхідні виправлення та повторно запускає програму.

Четвертий – виконується перевірка змінних на логічну узгодженість даних і допустимість значень у межах одного розділу й очищення даних за результатами запитів. Після внесення виправлень перевірка даних цього розділу проводиться повторно.

П'ятий – на цьому етапі перевіряються дані певного розділу на узгодженість із даними інших розділів та виправлення помилок за результатами запитів. Після внесення виправлень перевірка проводиться повторно.

Шостий – після очищення даних усіх розділів інструментарію формується файл скоригованого "Довідника граничних значень товарів та послуг" із рекомендованими значеннями змінних, обчисленими за реальними даними поточного звітного кварталу. Після коригування даних Довідника граничних значень процес контролю й очищення даних повторюється із самого початку.

Унесення змін у дані обстеження виконується тільки після підтвердження їх необхідності регіональними підрозділами ОУЖД за інформацією, уточненою в респондента. При цьому коригуватися дані можуть на будь-якому етапі процесу контролю та очищення даних.

Основні процедури очищення даних обстеження.

1). Етап підготовчих робіт.

- Створення копій оригінальних файлів даних, отриманих безпосередньо від головних управлінь статистики. Надалі вся робота проводиться тільки з файлами-копіями – робочими файлами, оригінальні ж файли зберігаються в первинному вигляді.

- Приведення кодованих змінних відповідно до встановлених кодифікаторів, наприклад кодифікатора товарів і послуг тощо.

- Приведення робочих файлів у вигляд, зручний для подальшої роботи:

- створюються нові змінні, необхідні для проведення аналізу за допомогою програми та наступної перевірки й очищення даних спеціалістом-економістом (наприклад, номер запису, змінні-прапорці, копії змінних, які найчастіше потребують очищення, тощо);

- додаються необхідні змінні з довідника (наприклад, коефіцієнти переведення в основні одиниці виміру, коди місць придбання товарів чи отримання послуг, мінімальні та максимальні значення кількості, вартості та ціни товару, норми споживання продуктів харчування на одну особу тощо).

2). Аналіз спеціалістами-економістами змісту коментарів фахівців з інтерв'ювання та внесення виправлень шляхом формування відповідних команд у синтаксис-файлах.

3). Змінні файлів даних (окрім змінних "код місця придбання", "код одиниці виміру", "кількість" та "вартість" товару / послуги) перевіряються на допустимість значень.

Для відображення наявності помилок по цих змінних використовується спеціальна змінна-індикатор наявності помилки. Якщо змінна-індикатор не дорівнює нулю, то значення певної змінної виходить за допустимі значення.

4). Перевірка значень змінних на узгодженість даних по першій і другій групах змінних.

- Змінні першої групи перевіряються на допустимість значень для цієї змінної (тобто питання з варіантами відповідей "так" або "ні" не можуть містити у відповідях вартісних показників).
- Змінні другої групи перевіряються на допустимість значень для цієї змінної (тобто питання, відповіді на які передбачають застосування вартісних показників, не можуть у відповідях містити показники інших одиниць виміру) .
- Перевіряється узгодженість змінних "номер ПТОВ" і "код регіону".
- Перевіряється узгодженість змінних "номер ПТОВ" і "тип населеного пункту".
- Для перевірки допустимих значень змінної "код регіону" та для перевірки відповідності значень змінних "номер ПТОВ", "код регіону" і "тип населеного пункту" використовується файл даних SPSS, який створюється на державному рівні.

5). Аналіз спеціалістами-економістами записів, де виявлені помилки, та внесення правок шляхом формування відповідних команд в синтаксис-файлах.

6). Перевірка значень змінних коду товару / послуги.

- Виконується перевірка на наявність пропущених і неіснуючих у довіднику граничних значень кодів, помилкові записи помічаються й вилучаються з подальшого аналізу в цьому сеансі запуску програми очищення даних.
- Відшукуються дублюючі записи у відповідних розділах щоденника поточних витрат і квартального запитальника, помічаються прапорцем, значення якого вказує на номер запису, який дублюється цим записом.

7). Аналіз спеціалістами-економістами записів, де виявлені помилки, та внесення правок шляхом формування відповідних команд у синтаксис-файлах. Перед унесенням виправлень необхідно:

- установити значення коду або з'ясувати, чому вказаний код не міститься в довіднику граничних значень;
- з'ясувати причини дублювання. Якщо це помилка введення даних у ПЕОМ, то запис вилучається з аналізу. Якщо ж причиною дублювання є однакові за характеристиками покупки чи послуги, здійснені чи отримані домогосподарством в один і той же день, то такі записи залишаються в обробці.

8). Перевірка значень змінних певного розділу статистичного формуляра.

- Будуються частотні розподіли для значень кожної змінної розділу та виконується перевірка на наявність пропущених і недопустимих значень змінних (кодовані змінні звіряються з довідниками). Помилкові записи помічаються й вилучаються з подальшого аналізу в цьому сеансі запуску програми очищення даних.
- Перевіряється логічна узгодженість даних розділу та їх значень. Помилкові записи помічаються та вилучаються з подальшого аналізу на цьому сеансі запуску програми.

9). Аналіз спеціалістами-економістами записів, де виявлені помилки, та внесення правок шляхом формування відповідних команд у синтаксис-файлах.

10). Після закінчення очищення даних усіх розділів статистичного формуляра виконується формування нового файла довідника граничних значень шляхом розрахунку нових мінімальних і максимальних значень показників за реальними даними поточного звітного кварталу. Значення цих показників визначається як 5% усічене середнє значення відповідного показника, визначеного за даними по всій вибірці респондентів. Структура файла сформованого довідника відповідає структурі файла діючого довідника.

11). По закінченні формування оновленого довідника поводитьься перевірка даних статистичного формуляра ще раз уже з урахуванням інформації оновленого довідника.

3.1.3. Імпутація відсутніх даних обстеження

При обробці даних ОУЖД застосовується процедура імпутації відсутніх значень. У додатку 4 зазначено, для яких частин статистичних формулярів проводиться імпутація відсутніх значень.

Запитальник основного інтерв'ю

При обробці даних запитальника основного інтерв'ю проводиться імпутація відсутніх даних за методом «найближчого сусіда», який полягає в пошуку найбільш подібного за характеристиками домогосподарства-донора та заповненні пропущених даних відповідними значеннями донора. Перед кожною процедурою імпутації формується декілька варіантів ключа подібності для імпутації відсутніх значень певного набору змінних. Перший ключ уключає максимально можливу кількість змінних. Якщо за першим ключем не вдається заповнити всі пропущені значення, використовується наступний ключ, який містить менший набір змінних.

При побудові ключа подібності для імпутації даних основного інтерв'ю домогосподарства групуються за основними характеристиками домогосподарства (наприклад, розмір домогосподарства, наявність пенсіонерів і дітей у їх складі тощо) у певному регіоні та місцевості. За відсутності домогосподарства-донора у певній місцевості та регіоні пошук донора відбувається в географічно близьких регіонах, де соціально-економічні умови життя населення наближені до місцевості, де мешкає реципієнт.

Щоденник поточних витрат домогосподарства

При обробці даних щоденника поточних витрат домогосподарств здійснюється імпутація відсутніх:

- днів заповнення щоденників;
- тижнів заповнення щоденників.

Із масиву первинних даних щоденників поточних витрат формується перелік домогосподарств, для яких відслідковуються незаповнені дні або тижні ведення щоденника. Для кожного домогосподарства окремо по кожному тижню ведення щоденника поточних витрат указується інформація щодо:

- причин незаповнення сторінки (розділу) щоденника;
- кількості заповнених днів по кожному розділу щоденника.

Зазначена інформація використовується для аналізу даних, які потребують імпутації, та відшукування запису-донора. Слід зазначити, що імпутація проводиться окремо по кожному розділу щоденника.

Незаповнений тиждень ведення щоденника імпутується даними іншого тижня по цьому ж домогосподарству. У випадках відсутності по домогосподарству даних по обох тижнях:

- 1) уточнюються дані щодо участі цього домогосподарства в обстеженні;
- 2) якщо домогосподарство брало участь, але з якихось причин не надало інформації щодо своїх витрат на обстежуваних тижнях, то за методом найближчого сусіда для імпутації відшуковується домогосподарство із подібними соціально-демографічними характеристиками. Для побудови ключа подібності використовуються дані основного інтерв'ю.

Для імпутації даних незаповнених днів ведення щоденника проводиться поступовий аналіз структури й обсягу витрат домогосподарства для визначення записів донорів:

- 1) по інших днях заповнення поточного тижня. Якщо спостерігається стандартний набір товарів і послуг, що їх споживає домогосподарство протягом усіх днів тижня, то дані заповнених днів розповсюджуються на незаповнений, при цьому використовується метод середнього значення;
- 2) по іншому тижню заповнення щоденника. Якщо спостерігається стандартний набір товарів і послуг протягом усіх днів тижня, то дані імпутуються середнім значенням заповнених днів іншого тижня заповнення. Можливий випадок, коли значення імпутуються даними відповідного дня з іншого тижня заповнення;

- 3) за даними домогосподарства за попередні періоди обстеження, тобто за попередній звітний квартал. Якщо структура витрат подібна для обох періодів, то імпутуються дані відповідного дня з попереднього звітного періоду;
- 4) відшуковується домогосподарство-донор з подібною або близькою структурою витрат і рівнем споживання та імпутуються дані відповідного дня, тобто використовується метод імпутації найближчого сусіда.

Описаний метод імпутації відсутніх днів і тижнів щоденника поточних витрат домогосподарств є досить трудомістким і потребує багато часу, оскільки для кожного випадку потрібен окремий підхід і немає детальної методики. Етап імпутації даних щоденників виконується протягом не менше *двох тижнів*.

У рідкісних випадках, коли виникає необхідність імпутації відсутніх значень щодо кількості, вартості або ціни окремого товару / послуги, використовуються методи заповнення відсутніх значень:

- найближчого сусіда – кількості спожитого товару / послуги;
- середнього значення по сукупності – для показників вартості та ціни товару / послуги;
- за пропорцією по цьому коду товару та певній сукупності – код місця придбання товару / послуги.

Перед формуванням у ГМУС у м. Києві масивів даних для передачі в Держкомстат виконується ще один етап обробки даних щоденників поточних витрат – розповсюдження даних двох тижневих щоденників на весь квартал обстеження за допомогою системи коефіцієнтів. Цю систему коефіцієнтів формує Держкомстат окремо по кожному розділу щоденника поточних витрат та окремо по товарах і послугах і передає ГМУС у м. Києві. Більшість показників поширюються із застосуванням коефіцієнта 6,5 (13 тижнів кварталу, поділених на 2 тижні ведення щоденників поточних витрат), для визначення решти застосовуються експертні оцінки на основі частоти покупок (витрат) конкретних товарів / послуг.

Квартальний запитальник про витрати і доходи домогосподарств

При обробці даних запитальника квартального інтерв'ю також проводиться імпутація відсутніх даних.

Для заповнення відсутніх даних стосовно витрат на житлово-комунальні послуги використовується метод пошуку найближчого сусіда. Формування ключів подібності для пошуку домогосподарств-донорів здійснюється з використанням даних основного інтерв'ю щодо соціально-демографічних характеристик домогосподарств. Файл із даними основного інтерв'ю після завершення обробки даних надходить у відповідний підрозділ ГМУС у м. Києві від департаменту обстежень домогосподарств Держкомстату.

Перший ключ формується за максимально можливою кількістю змінних характеристик домогосподарства. Якщо за першим ключем не вдається заповнити всі пропущені значення, використовується наступний ключ, який містить менший набір змінних. Домогосподарства-донори відшуковуються за типом місцевості та регіоном проживання, якщо таких не існує, то пошук проводиться у територіально близьких регіонах.

Імпутація відсутніх даних по житлово-комунальних послугах для домогосподарств, які проживають у гуртожитках, проводиться методом середнього (заповнення середнім значенням). При використанні цього методу розраховується середнє значення показника по групі домогосподарств із певними характеристиками. У цьому випадку середнє значення розраховується групі домогосподарств, які:

- 1) проживають в одному гуртожитку;
- 2) проживають в одній місцевості та отримують ідентичний набір послуг;
- 3) проживають у територіально близьких регіонах і отримують ідентичний набір послуг.

3.1.4. Файли даних та їх зберігання

За результатами етапів обробки даних, які виконуються у ГМУС у м. Києві, формуються та передаються у Держкомстат файли даних по статистичних формулярах ОУЖД:

- основного інтерв'ю та контрольної картки по домогосподарствах і по їх членах;
- щоденників поточних витрат домогосподарств;
- квартального інтерв'ю та контрольної картки з урахуванням змін у складі домогосподарства;
- одноразових анкет (передаються одночасно із файлами даних відповідного квартального інтерв'ю).

Дані для кожного року обстеження зберігаються на сервері в окремій БД, яка після вивантаження в статистичний пакет SPSS архівується, записується на зовнішній носій, а потім вилучається. У БД зберігається інформація, необхідна лише для поточної обробки річних даних. Вона поділяється на два види:

перший – дані, що використовуються протягом усього обстеження та поновлюються в процесі обробки даних:

- нормативно-довідкова інформація (усі довідники),
- каталог ПТОВ,
- каталог фахівців з інтерв'ювання,
- каталог домогосподарств,
- каталог членів домогосподарств (дані по контрольних картках).

другий – дані основного, квартального інтерв'ю, щоденники поточних витрат, анкети.

3.2. Функції та завдання, що виконуються у Держкомстаті

Роботи з обробки даних ОУЖД, які виконує департамент обстежень домогосподарств Держкомстату (надалі – департамент), передбачають реалізацію таких етапів:

- логічний контроль і перевірка узгодженості даних;
- формування файлів даних аналізу за результатами ОУЖД за квартал ;
- формування системи статистичних ваг домогосподарств та їх членів для розповсюдження даних обстеження на генеральну сукупність;
- аналіз зіставності/порівнянності та узгодженості/последовності результатів обстеження зі статистичною інформацією, що виробляється іншими підрозділами Держкомстату;
- розрахунок характеристик надійності основних показників доходів і витрат домогосподарств;
- підготовка інформаційно-аналітичних матеріалів за результатами обстеження: аналітичних таблиць, бюлетенів, збірників і мікрофайлів даних за кожен квартал, перше півріччя, дев'ять місяців та за рік у цілому, а також підготовка звітів із якості результатів обстеження.

Обробка даних у департаменті здійснюється відповідно до Технологічного плану державних статистичних спостережень на поточний рік (див. додаток 5). Роботи виконуються на основі файлів даних, отриманих від ГМУС у м. Києві. У департаменті здійснюється обробка даних усіх типів статистичних формулярів ОУЖД:

- дані одноразових обстежень обробляються разом із даними запитальників квартального інтерв'ю (як один із розділів запитальника);
- дані щоденників поточних витрат домогосподарств і квартальних запитальників об'єднуються й надалі обробляються разом.

Усі етапи обробки даних проводяться в програмному середовищі SPSS та пов'язаних програмних середовищах (G-Calib – калібрація системи статистичних ваг, WESVAR – розрахунок характеристик надійності показників, Excel – формування інформаційно-аналітичних матеріалів за результатами ОУЖД).

3.2.1. Логічний контроль і узгодження файлів даних

Запитальник основного інтерв'ю

До департаменту з ГМУС у м. Києві надходять 4 файли з результатами основного інтерв'ю (див. таблицю 3). При обробці даних запитальників основного інтерв'ю використовується декілька логічних контролів:

- перевірка відповідності кодів ПТОВ каталогу ПТОВ;
- перевірка відповідності типів населеного пункту каталогу ПТОВ;
- логічний контроль узгодженості даних щодо рівня освіти членів домогосподарств (наприклад, дитині на момент проведення інтерв'ю виповнилося 8 років, а інформація щодо її рівня освіти відсутня), житлових умов і характеристик земельних ділянок домогосподарств.

При виникненні неузгодженості даних надсилається запит щодо некоректних даних у ГМУС у м. Києві. Після виправлення всіх неузгодженостей, проведення повторних контролів здійснюються розрахунки аналітичних показників (показники щодо домогосподарств, у складі яких є діти, які не мають одного чи обох батьків, домогосподарств, у складі яких є молоді сім'ї і т.д.) та формуються два файли аналізу даних за результатами обробки запитальників основного інтерв'ю щодо соціально-демографічних характеристик респондентів:

- 1) домогосподарств;
- 2) членів домогосподарств.

Квартальний цикл ОУЖД

Щокварталу до департаменту надходять:

- 7 файлів із даними щоденників поточних витрат домогосподарств (відповідно до кількості розділів щоденника);
- 10 файлів із даними квартального запитальника про витрати і доходи домогосподарств;
- файл даних щодо витрат домогосподарствами на будівництво та ремонт (анкета №1);
- у відповідні квартали додаються файли з даними одноразових опитувань (анкета №2 й анкета №3);
- файл з інформацією, необхідною для визначення рівня участі домогосподарств в опитуванні та причин відмови від участі;
- файл з актуалізованою інформацією щодо членів домогосподарств, які брали участь в обстеженні.

Усі файли даних у назві містять інформацію щодо періоду обстеження (рік і квартал обстеження). Квартальна обробка даних здійснюється за допомогою спеціально розроблених синтаксис-програм SPSS. У департаменті організована локальна мережа комп'ютерів (з обмеженим доступом визначеного кола працівників), яка створена з метою збереження повної конфіденційності інформації щодо результатів обстеження умов життя домогосподарств.

Для автоматизованої обробки квартальних даних ОУЖД розроблена спеціальна структура розташування вхідних, робочих (розрахункових) та кінцевих файлів даних обстеження. На ПЕОМ, де ведеться обробка даних, створені спеціальні директорії за такою структурою. На початку року обстеження створюється головна директорія [**Processing_20XX**] (де XX – 2 цифри року), яка містить у собі такі директорії:

- 1) [Original] – наприкінці року в цій директорії знаходяться чотири директорії відповідно до кожного кварталу, які містять у собі файли даних, що надійшли з ГМУС у м. Києві.
- 2) [Temporary] – дані в цій директорії зберігаються лише протягом обробки одного кварталу, оскільки в цій директорії зберігаються проміжні та розрахункові файли за результатами контролю, агрегування та спеціальних розрахунків. Після завершення обробки даних одного квартального інтерв'ю дані, які містяться в цій папці, архівуються та зберігаються на CD-диску. Папка повністю звільняється від інформації для обробки даних наступного квартального інтерв'ю.
- 3) [Calcul_files] – наприкінці року в цій директорії знаходяться чотири директорії відповідно до кожного кварталу та окремі директорії для даних за півріччя, 9 місяців та рік. У цій директорії зберігаються файли аналізу даних по домогосподарствах і по членах домогосподарств, а також файли даних окремо по кожному розділу щоденника та квартального запитальника.
- 4) [SPSS_FILES_XKV] – у цій директорії зберігаються всі алгоритми розрахунків, які використано під час обробки квартальних даних.

5) [TABLES_for_KV] – наприкінці року в цій директорії знаходиться чотири директорії відповідно до кожного кварталу й окремі директорії для даних за півріччя, 9 місяців та рік. У цій директорії зберігаються Excel-файли розрахунків та файли для друку публікацій.

Щоденник поточних витрат домогосподарства

У департаменті схема обробки даних для всіх розділів щоденників однакова, за винятком першого та четвертого розділів (які містять інформацію щодо витрат домогосподарств на продукти харчування та спожиті продукти, отримані з особистого підсобного господарства та з інших джерел), які розробляються за більш розширеною схемою.

Для кожного розділу щоденника поточних витрат після отримання з ГМУС у м. Києві здійснюється повторний контроль, під час якого проводиться більш детальні перевірки логічної узгодженості даних. Окрім цього, виконується контроль «входу-виходу», тобто перевірка відповідності та збереження даних перед процедурою та після її виконання, що дозволяє не втрачати частину інформації під час обробки даних.

Проводиться виявлення кодів товарів, які за методологією обстеження не можуть бути присутніми у певному розділі щоденника. Наприклад, у п'ятому розділі щоденника поточних витрат, у якому міститься інформація щодо витрат домогосподарств лише на непродуктивні товари, не може бути кодів, які відносяться до категорії послуг. Перевіряється правильність кодування місця придбання товару. Наприклад, по коду товару «ремонт та технічне обслуговування великих домашніх приладів та електро побутової техніки (крім встановлення)» за методологією не може бути зазначений код місця придбання товару «2 – на ринку». У випадку встановлення будь-якого з вищенаведених неузгоджень робиться запит, і якщо виявляється помилка, вона виправляється у ГМУС у м. Києві на основі уточненої в домогосподарстві інформації, а до департаменту передається виправлений файл по відповідному розділу щоденника поточних витрат. Якщо під час контролю даних помилок не виявлено, у файлі виводу з'являється повідомлення: «Помилки не виявлено» та «Кількість записів правильна».

Наступним кроком при обробці даних щоденників поточних витрат є приведення даних обстеження щодо витрат домогосподарств методом групувань відповідно до КІСЦ (класифікації індивідуального споживання товарів та послуг за цілями, що відповідає міжнародній класифікації COICOP-HBS)) до рівня категорій (п'ятизначний код).

Після виконання зазначеної процедури проводяться контролю:

- «входу-виходу»;
- правильності розрахованих показників за місцем придбання, наприклад, сума вартісних показників по чотирьох напрямках придбання товару (у магазині, на ринку, зі складу підприємств і в громадян) повинна дорівнювати загальній вартості товару.

Для першого та четвертого розділів щоденника поточних витрат виконуються додаткові розрахунки, які стосуються визначення обсягів споживання продуктів харчування домогосподарствами у натуральному вираженні за основними групами (хліб і хлібопродукти, м'ясо та м'ясопродукти, молоко та молокопродукти та ін.)

Для здійснення цих розрахунків створюється додатковий файл даних, де міститься інформація щодо коефіцієнтів перерахунку в основний продукт по кожному продовольчому коду товару, а також вміст білків, жирів, вуглеводів і калорійність кожного продукту. Цей файл даних додається до файлів даних першого та четвертого розділів щоденника поточних витрат. Результатом є файли даних першого та четвертого розділів щоденника поточних витрат з обсягами споживання по кожному домогосподарству для двадцяти груп продуктів харчування з групуванням за місцем їх придбання, а також із характеристиками якості раціону спожитих респондентами продуктів. На цьому етапі також застосовується контроль правильності розрахованих показників за місцем придбання.

Квартальний запитальник про витрати і доходи домогосподарства

Після завершення обробки даних щоденників поточних витрат виконується обробка файлів даних квартальних запитальників. За своєю структурою файли даних квартального запитальника про витрати і доходи домогосподарств для розділів 1, 2.1, 2.2, 2.3 та 2.4

відповідають структурі файлів даних розділів щоденника поточних витрат. Обробка зазначених розділів квартального запитальника здійснюється за структурою обробки розділів щоденника (як описано вище). Відповідно перший розділ квартального запитальника, який містить інформацію щодо витрат домогосподарств на купівлю продуктів харчування для тривалого користування, обробляється за схемою обробки першого розділу щоденника поточних витрат.

Для кожного з розділів 3, 4, 5, 6, 8 і анкети №1 (у відповідні квартали анкети №2 та анкети №3) квартального запитальника окремо проводиться контроль на мінімальне та максимальне значення показників. Необхідність проведення цього контролю повторно зумовлюється можливістю псування або втрати даних при передачі з ГМУС у м. Києві. У випадку виявлення помилок, вони виправляються у ГМУС у м. Києві на основі уточненої в домогосподарстві інформації.

Після проходження контролів зазначені файли даних зберігаються у директорії [Temporary], а файли-оригінали залишаються без змін у директорії [Original]. Надалі вся робота ведеться лише зі збереженими копіями файлів даних.

Наступним кроком є об'єднання файлів розділів 3, 4, 5, 8 і анкети №1 в один, що дозволяє в подальшому здійснити контроль на взаємоузгодження показників між окремими розділами квартального запитальника.

По завершенні всіх підготовчих робіт з файлами даних квартального запитальника розділів 3, 4, 5, 8 і анкети №1 виконується більш детальна перевірка на взаємоузгодженість даних цих розділів за допомогою спеціально розробленої синтаксис-програми, що містить перевірку всіх можливих варіантів неузгодженості між окремими змінними та розділами.

У разі виявлення будь-яких помилок при проходженні цього контролю робиться запит до ГМУС у м. Києві, де здійснюється в установленому порядку виправлення цих помилок і передача до департаменту нового файла даних по конкретному розділу, де були виявлені помилки.

Потім для окремих товарів/послуг розраховуються чотири показники щодо місця придбання:

- вартість придбання в магазині;
- вартість придбання на ринку;
- вартість придбання зі складу підприємства;
- вартість придбання в громадян.

По завершенні розрахунків по кожній змінній проводиться арифметичний контроль правильності розподілу за місцем придбання.

Файли даних щоденника поточних витрат і квартального запитальника мають різну структуру:

- у файлі даних щоденника надається інформація щодо коду домогосподарства, коду товару та вартості по кожному коду товару;
- файл даних квартального запитальника містить окремі змінні, які характеризують витрати домогосподарств на різні товари та послуги.

Основна мета наступного етапу обробки даних – отримання зведених витрат домогосподарств за квартал (з обох статистичних формулярів) в одному файлі за структурою агрегації категорій КІСЦ (п'ятизначний код). Для цього структура файла даних квартального запитальника приводиться у відповідність із структурою файла щоденника поточних витрат. Для виконання цієї процедури створено спеціальну синтаксис-програму з використанням макросу.

У результаті кожній змінній у файлі квартального запитальника присвоюється відповідний код товару чи послуги з одночасним розподілом за місцем придбання, тобто формується файл зі структурою щоденника, що вміщує дані квартального запитальника. Для перевірки правильності розрахунків наступним кроком виконується контроль «входу-виходу». Після завершення отриманий файл зі структурою щоденника обробляється за схемою обробки файлів щоденника поточних витрат, яка була описана вище.

Для визначення витрат домогосподарств за більш детальним розподілом товарів/послуг, ніж категорій КІСЦ, здійснюється розрахунок витрат складових частин категорій КІСЦ:

- Кожен розділ щоденника поточних витрат і квартального запитальника обробляється окремо.
- До файла даних додається спеціальний файл, у якому міститься змінна, що визначає відповідність того чи іншого коду товару певній категорії КІСЦ.
- На основі цих даних розраховуються показники складових категорії КІСЦ з урахуванням місця придбання товару/послуги.
- Виконується контроль «входу-виходу» роботи цієї процедури.

У результаті проведених розрахунків формуються окремі файли даних по 1, 2, 3, 5, 6 розділах щоденника поточних витрат та 1, 2.1, 2.3, 2.4, 3 – 5, 9 розділах квартального запитальника (9 розділ – анкета №1), приведених у відповідність із структурою щоденника поточних витрат. Технологія обробки даних щоденника та квартального запитальника для цих розділів однакова.

Файли даних зберігаються на ПЕОМ, де проводились розрахунки, також робиться копія на ПЕОМ, де ведеться квартальна обробка.

3.2.2. Формування файлів аналізу даних

При формуванні файла аналізу даних щодо витрат та ресурсів домогосподарств першим формується файл загальних витрат.

Розрахунок загальних показників витрат домогосподарств на непродовольчі товари та послуги передбачає такі етапи:

- 1) об'єднання в єдиний файл даних розділів 2, 3, 5, 6 щоденника поточних витрат і розділів 2.1, 2.2, 2.3, 2.4 квартального запитальника;
- 2) агрегування даних цього файла за кодом товару/послуги з одночасним додаванням міток (стислих описів змісту) змінних і здійснення контролю «входу-виходу»;
- 3) приєднання до вищезазначеного файла даних по 3 – 5, 9 розділах квартального запитальника, приведеного у відповідність із структурою щоденника;
- 4) розрахунок загальних показників витрат домогосподарств на непродовольчі товари і послуги з розподілом за місцем придбання та додавання міток змінних.

Для розрахунку загальних показників витрат домогосподарств на продукти харчування здійснюються такі етапи:

- 1) об'єднуються файли даних перших розділів щоденника поточних витрат і квартального запитальника;
- 2) дані агрегуються за кодом товару та додаються мітки змінних;
- 3) здійснюється контроль «входу-виходу».

Отриманий файл даних містить показники витрат домогосподарств відповідно до КІСЦ з розподілом показників за місцем придбання товару.

Агреговані файли даних витрат домогосподарств на продукти харчування та на непродовольчі витрати та послуги об'єднуються. На основі зведеного файла розраховуються показники витрат на рівні груп, класів і категорій КІСЦ.

Окремо здійснюються розрахунки по показниках доходів домогосподарств за такою схемою:

- на основі файла даних розділу 6 квартального запитальника по кожному члену домогосподарства формується файл даних індивідуальних доходів домогосподарств;
- дані розділу 6 агрегуються по кодах домогосподарств і об'єднуються з даними розділу 8, який містить інформацію щодо доходів домогосподарства в цілому;
- на основі отриманого файла розраховуються загальні показники за видами доходів домогосподарств у цілому;
- у зведеному файлі на основі даних 6 і 8 розділів квартального запитальника здійснюються розрахунки щодо обсягів пільг і субсидій, які отримують домогосподарства.

Для отримання загального файла даних з інформацією як про витрати, так і про доходи домогосподарств об'єднуються файли даних щодо:

- соціально-демографічних характеристик домогосподарств;
- витрат домогосподарств на товари та послуги відповідно до КІСЦ;

- отриманих домогосподарствами доходів;
- щоденників поточних витрат щодо спожитих продуктів харчування, отриманих з особистого підсобного господарства та з інших джерел.

На основі об'єднаного файла розраховуються показники доходів і витрат домогосподарств:

- споживчі грошові витрати;
- грошові витрати;
- сукупні витрати;
- грошовий дохід;
- загальний дохід;
- сукупні ресурси.

Файли даних, створені за результатами розрахунків витрат рівня категорій КІСЦ (п'ятизначний код), додаються до зведеного файлу даних витрат і доходів домогосподарств.

Така схема розрахунку дозволяє здійснити додатковий контроль. Оскільки розрахунки по рівнях розділів, груп, класів та по рівню категорій КІСЦ ведуться за різною схемою та відокремлені один від одного (на різних ПЕОМ), це дозволяє проконтролювати правильність виконання розрахунків. Наприклад, сума показників рівня категорій повинна дорівнювати значенню суми показників рівня класів і т.д. Якщо суми не збігаються, то виконуються розрахунки як по рівнях розділів, груп, класів КІСЦ, так і по рівню категорій, після чого повторно проводиться контроль.

У результаті наведених процедур формується перший зведений файл аналізу даних витрат та доходів домогосподарств, який, зокрема, містить інформацію щодо витрат на рівні розділів, груп, класів і категорій КІСЦ.

Одночасно з обробкою даних квартального запитальника відбувається підготовка файлу даних для розрахунку квартальних статистичних ваг, які використовуються для поширення даних вибіркового обстеження на генеральну сукупність. На основі файлів даних щодо соціально-демографічних характеристик здійснюються розрахунки показників, які беруть участь в розрахунках системи статистичних ваг (визначаються типи домогосподарств залежно від наявності дітей у його складі, від наявності дорослих осіб у їх складі, від наявності працюючих осіб у складі домогосподарства, розраховуються показники віку членів домогосподарств – респондентів та ін.). Процедура розрахунку системи статистичних ваг описана у пункті 3.2.3.

Подальша обробка й аналіз даних відбувається з урахуванням системи статистичних ваг домогосподарств і членів домогосподарств. Остаточний файл аналізу даних щодо витрат і ресурсів домогосподарств, який є основою для формування таблиць публікацій за результатами обстеження, формується на основі зведеного файлу даних доходів і витрат домогосподарств і файлу з результатами виконання процедури формування системи статистичних ваг даних ОУЖД.

За допомогою спеціальної синтаксис-програми здійснюються розрахунки для включення до файлів аналізу додаткових основних показників, необхідних для подальших розрахунків аналітичних таблиць, які входять до складу публікації:

- Належність домогосподарства до однієї з типологічних груп, розподілу домогосподарств за такими ознаками:
 - наявність дітей;
 - кількість осіб у домогосподарстві;
 - наявність працюючих осіб у домогосподарстві;
 - за регіонами,
 - за економічними районами та за іншими ознаками.
- Належність домогосподарства до групи за рівнем середньодушових грошових та загальних доходів:
 - із середньодушовими доходами нижче прожиткового мінімуму;
 - із середньодушовими доходами нижче середнього рівня;
 - із середньодушовими доходами нижче 50% середнього рівня доходів;
 - із середньодушовими доходами нижче медіанного рівня доходів;

- із середньодушовими доходами 50% медіанного рівня;
- із середньодушовими доходами 60% медіанного рівня;
- із середньодушовими доходами 75% медіанного рівня доходів.
- Належність домогосподарства до певної децильної групи розподілу домогосподарств за рівнем середньодушових грошових і загальних доходів.

Також до файла аналізу даних додається окремо сформований файл щодо обсягу споживання домогосподарствами продуктів харчування. Файл містить інформацію по кожному домогосподарству щодо споживання продуктів харчування за групами:

- Хлібні продукти (хліб та макаронні вироби у перерахунку на борошно; крупи, бобові)
- Картопля та картоплепродукти
- Овочі
- Плоди, ягоди, горіхи та виноград
- Баштанні
- Гриби
- Цукор та кондитерські вироби
- Олія, маргарин та інші жири
- Мед
- М'ясо та м'ясопродукти, включаючи субпродукти і жир
- Молоко та молочні продукти
- Яйця та яйцепродукти
- Риба та рибопродукти
- М'ясо та м'ясопродукти (без сала)
- Молоко та сир
- Масло
- Сало та жири тваринні
- Олія та інші рослинні жири
- Фрукти, ягоди, горіхи, виноград, кавуни, дині
- Овочі, картопля, гриби.

Файл аналізу даних по членах домогосподарств формується на основі файла даних розділу 6 квартального запитальника по кожному члену домогосподарства. До цього файла даних додається інформація щодо соціально-демографічних характеристик домогосподарства по кожному члену домогосподарства, а саме: показників щодо статі, віку, ступеня родинних стосунків з головою домогосподарства, рівня освіти, соціально-економічного статусу кожного члена домогосподарства. На основі цього файла даних також здійснюються розрахунки щодо індивідуальних доходів кожного члена домогосподарства.

Дані одноразових опитувань домогосподарств обробляються в порядку, аналогічному даним квартального запитальника. Обробка даних цих анкет побудована за єдиною схемою: контроль вхідних даних, розрахунок показників для таблиць публікацій та їх уключення до загального файла аналізу даних, тобто дані по кожному члену домогосподарства додаються до файла аналізу по членах домогосподарства, дані по домогосподарствах – до файла аналізу по домогосподарствах. Після об'єднання даних основних файлів аналізу з даними анкет отримуємо кінцеві файли аналізу даних за результатами вибіркового обстеження умов життя домогосподарств.

Файли аналізу по домогосподарствах і членах домогосподарств готуються не тільки за квартал, а й за півріччя, 9 місяців та рік. При цьому розрахунки проводяться за колом домогосподарств, які брали участь в ОУЖД протягом відповідних звітних періодів (тобто ті респонденти, які відмовилися від участі у II кварталі, виключаються з розрахунків показників за півріччя, 9 місяців та за рік).

Для користувачів інформації за даними ОУЖД формуються спеціальні файли даних – мікрофайли:

- по домогосподарствах;
- по членах домогосподарств.

Мікрофайли по домогосподарствах містять по кожному домогосподарству основні показники, що розраховуються за результатами ОУЖД, зокрема щодо:

- соціально-демографічних характеристик домогосподарств,
- витрат домогосподарств на товари та послуги відповідно до КІСЦ на рівнях розділів, груп, класів та категорій,
- доходів,
- споживання домогосподарствами продуктів харчування,
- доступності для домогосподарств медичних послуг,
- наявності в домогосподарствах товарів тривалого користування,
- самооцінки домогосподарствами рівня своїх доходів та ін. дані одноразових опитувань.

У файлі аналізу даних по членах домогосподарств міститься інформація щодо:

- соціально-демографічних характеристик членів домогосподарств,
- індивідуальних доходів членів домогосподарств,
- стану здоров'я членів домогосподарства,
- народжуваності дітей жінками,
- питань тютюнокуріння тощо.

Для зручності користування для кожного з мікрофайлів формуються описи структури файлів, що містять таку інформацію:

- назву змінної;
- опис змінної – розшифровується назва та додаються коментарі до змінної. Для показників витрат на товари / послуги в описі змінної наводиться повна назва статті витрат відповідно до КІСЦ;
- значення відповіді – надається перелік можливих значень.

Описи надаються користувачам разом із мікрофайлами даних ОУЖД.

3.2.3. Формування системи статистичних ваг

Одним із етапів обробки є процедура розповсюдження даних ОУЖД на генеральну сукупність шляхом розрахунку системи статистичних ваг. Для розрахунку оцінок показників, що репрезентують (представляють) усе населення (генеральну сукупність), в ОУЖД обчислюється система статистичних ваг, яка враховує:

- 1) загальну ймовірність відбору домогосподарств – ваги плану (дизайну) вибірки;
- 2) фактичні рівні участі домогосподарств у обстеженні – ваги з урахуванням відмов від обстеження;
- 3) додаткову зовнішню інформацію щодо соціально-демографічних характеристик домогосподарств і їх членів – калібраційні ваги. Окрім того, система ваг є взаємоузгодженою на рівні домогосподарств і рівні осіб.

Система статистичних ваг формується за загальноприйнятими принципами та статистичними підходами:

- 1) розрахунок ваг плану (дизайну) вибірки здійснюється в середовищі Excel на початку дії територіальної вибірки – *раз на n'ять років*;
- 2) коригування ваг плану (дизайну) вибірки виконується в середовищі SPSS під час обробки даних основного та квартальних інтерв'ю – *n'ять разів на рік*;
- 3) процедура калібрації ваг реалізована на основі рекомендацій міжнародних консультантів з використанням спеціально розроблених і стандартних програмних засобів, зокрема за допомогою програми g-Calib. Ця процедура застосовується для даних основного та квартальних інтерв'ю (в тому числі для компенсації відмов респондентів від квартальних інтерв'ю) – *n'ять разів на рік*.

Статистичні ваги одиниці вибірки (домогосподарства) або ваги плану (дизайну) вибірки визначаються як обернена величина до загальної ймовірності відбору цієї одиниці, яка включає всі три ймовірності відбору територій і домогосподарств:

$$w_{Bi} = \frac{1}{p_{1i} \cdot p_{2i} \cdot p_{3i}},$$

де P_{1i} – імовірність відбору одиниць першого ступеня (ПТОВ). Одиницями відбору першого ступеня є окремі міста (міськради) та селища міського типу в міській місцевості та сільські ради – у сільській місцевості;

P_{2i} – імовірність відбору одиниць другого ступеня (ВТОВ). У міських поселеннях – це (на прикладі циклу 2009-2013 рр.) виборчі дільниці, створені для проведення Позачергових виборів народних депутатів України 30 вересня 2007р. (або їх сукупності). Для сільської місцевості P_{2i} – це теоретична ймовірність відбору домогосподарств у межах відібраних ПТОВ (сільських рад);

P_{3i} – теоретична ймовірність відбору домогосподарств у межах ВТОВ у міській місцевості;

Урахування рівнів відмов домогосподарств від участі в ОУЖД відбувається шляхом коригування ваг плану (дизайну) на реальний рівень участі. Коригування системи ваг здійснюється за допомогою спеціальної системи комірок (класів зважування), які формуються за класифікаційними змінними:

- економічний район – 8 економічних районів (економічне районування розроблене Радою по вивченню продуктивних сил України НАН України для використання в програмі “Україна – 2010”);
- тип місцевості – 3 типи (“велике” місто, “мале” місто та село);
- група за кількістю осіб у домогосподарстві – 4 групи (1 особа, 2 особи, 3 особи, 4 і більше осіб).

Ця процедура коригування ваг включає такі кроки:

1) розподілення вибіркової сукупності домогосподарств на групи (комірки зважування), що охоплюють усю сукупність домогосподарств та не перетинаються – щоразу контролюється достатня наповнюваність комірок (не менше 20);

2) обчислення рівня участі одиниць в обстеженні для кожної комірки (на основі файлів даних основного та квартальних інтерв'ю (усі домогосподарства, що були відібрані та домогосподарства, що взяли участь в опитуванні));

3) визначення єдиного коефіцієнта коригування ваг для всіх одиниць конкретної комірки як величини, оберненої до рівня участі домогосподарств в обстеженні;

4) коригування ваг плану (дизайну) здійснюється у файлі даних SOI_DGXX.sav шляхом множення статистичних ваг на коефіцієнт коригування.

Для підвищення рівня надійності оцінок показників і узгодження результатів вибіркового обстеження із зовнішньою інформацією до статистичних ваг застосовується процедура калібрації. В ОУЖД коригування системи статистичних ваг здійснюється з метою узгодження певних показників обстеження з інформацією двох зовнішніх джерел:

- 1) даними демографічної статистики станом на початок року обстеження щодо:
 - чисельності постійного населення за регіонами та типами місцевості;
 - статеві-вікової структури населення за регіонами для вікових груп:
 - жінки у віці до 18 років;
 - жінки у віці 18-35 років;
 - жінки у віці 36-54 роки;
 - жінки у непрацездатному віці;
 - чоловіки у віці до 18 років;
 - чоловіки у віці 18-35 років;
 - чоловіки у віці 36-59 років;
 - чоловіки у непрацездатному віці.
- 2) даними статистичної звітності щодо населення, яке знаходиться в інституційних закладах.

На основі процедури калібрації визначаються вагові коефіцієнти, що забезпечують відповідність результатів обстеження зовнішнім даним та мінімізують відхилення ваг до та після калібрації. Застосування методу калібрації також вирішує задачу узгодження ваг рівня домогосподарств і рівня осіб при відсутності зовнішньої інформації, тобто ваги рівня осіб розглядаються як зовнішня інформація для уточнення ваг домогосподарств. У результаті калібрації статистичних ваг отримується єдина система ваг для домогосподарств і осіб.

Для здійснення процедури калібрації за допомогою програми g-Calib 2.0 готується файл даних SPSS, який містить необхідну для калібрації інформацію по кожному домогосподарству:

– змінні ідентифікатори:

код домогосподарства (code_fam);

код регіону (cod_obl);

тип населеного пункту (tp_ns_p);

– статистична вага домогосподарств до калібрації (wgt_kv):

для даних основного інтерв'ю – це ваги з урахуванням дизайну вибірки та відмов від інтерв'ю;

для даних I кварталу – це ваги після калібрації основного інтерв'ю;

для даних II-IV кварталу – це вагу після калібрації попереднього кварталу.

– змінні-характеристики домогосподарства, по яких здійснюється узгодження результатів обстеження із зовнішньою інформацією:

розмір домогосподарства (hsize);

ознака наявності у домогосподарстві дітей (dity);

кількість у домогосподарстві чоловіків віком до 17 років (m_0_17);

кількість у домогосподарстві чоловіків віком 18-35 років (m_18_35);

кількість у домогосподарстві чоловіків віком 36-59 років (m_36_59);

кількість у домогосподарстві чоловіків віком 60 років і старше (m_60);

кількість у домогосподарстві жінок віком до 17 років (w_0_17);

кількість у домогосподарстві жінок віком 18-35 років (w_18_35);

кількість у домогосподарстві жінок віком 36-54 роки (w_36_54);

кількість у домогосподарстві жінок віком 55 років і старше (w_55);

– спеціальні змінні, які необхідні для роботи програми g-Calib 2.0:

номер страти (no_strt) – регіони України, які послідовно нумеруються від 1 до 27 (1 – Автономна Республіка Крим, ..., 27 – м. Севастополь);

спеціальний параметр калібрації (nullwei) – приймає значення 1.

У діалогових вікнах програми g-Calib задаються:

– основні параметри налаштування роботи процедури;

– задаються умови задачі калібрації та зовнішня інформація по кожному розрізу розрахунку ваг;

– визначається метод калібрації.

Результат роботи програми зберігається у файлі даних SPSS U_all.svo, в якому по кожному домогосподарству надається:

– інформація із вхідного файла даних:

– змінні, які визначали умови задачі калібрації, фактично коефіцієнти калібрації:

– змінні, отримані в процесі розрахунку:

параметр масштабу калібрації (scale);

початкові ваги (scawei);

ваги після калібрації (calwei);

величина зміни ваг (g_wei).

Детальні коментарі роботи процедури калібрації статистичних ваг записуються у файлі виводу temp.spo. Результати розрахунків містяться у файлі результатів розрахунку.

Сформована система статистичних ваг приєднуються до файла аналізу даних та використовуються на наступних етапах обробки й аналізу даних ОУЖД.

3.2.4. Аналіз зіставності/порівнянності та узгодженості/послідовності результатів обстеження

При обробці даних за результатами ОУЖД важливим етапом є порівняння отриманих даних з аналогічними показниками, отриманими за результатами суцільних (невибіркових) спостережень, звітності тощо.

I. При формуванні файлів аналізу з результатами основного інтерв'ю проводяться порівняння:

- даних щодо статево-вікової структури населення з аналогічними даними демографічної статистики,
- даних щодо середньої площі земельної ділянки домогосподарства з аналогічними даними статистики сільського господарства,
- даних щодо середнього розміру загальної площі житла на одну особу з аналогічними даними соціальної статистики.

II. При формуванні файлів аналізу як за квартал, так і за рік у цілому здійснюються порівняння:

- основних зведених показників (заробітна плата, грошовий дохід, сукупні ресурси, грошові витрати, сукупні витрати в цілому та на рівні розділів КІСЦ) з даними національних рахунків,
- показників щодо споживання домогосподарствами продуктів харчування з аналогічними даними статистики сільського господарства,
- показників щодо окремих пільг і субсидій на оплату житла, комунальних продуктів та послуг з аналогічними даними соціальної статистики,
- показників щодо кількості користувачів послугами Інтернету удома з аналогічними показником за даними соціальної статистики.

Одночасно здійснюється перевірка даних кінцевого файла аналізу ОУЖД.

- Відповідно до структури файла даних формується спеціальна Excel-таблиця, яка містить дані щодо витрат домогосподарств на рівні розділів, груп, класів та категорій КІСЦ, також окремими блоками формуються показники складових доходів домогосподарства та всі загальні агреговані показники (грошовий дохід, загальний дохід, сукупні ресурси, грошові та сукупні витрати).
- Усі показники поточного звітного періоду порівнюються в динаміці з відповідними показниками попереднього звітного періоду та показниками аналогічного періоду минулого звітного року.

У випадках суттєвих розходжень проводиться поглиблений аналіз можливих зміщень, похибок або інших причин таких ситуацій.

3.2.5. Розрахунки характеристик надійності оцінок показників

При аналізі даних, отриманих за результатами вибіркового обстеження, вирішальне значення для інтерпретації та подальшого аналізу результатів має інформація щодо їх надійності. Інформація з недостатньою надійністю не може бути використана для адекватного аналізу результатів обстеження.

Для оцінки надійності показників ОУЖД використовуються такі характеристики:

- стандартна похибка вибірки;
- дисперсія вибірових оцінок;
- гранична похибка вибірки;
- відносна стандартна похибка вибірки або коефіцієнт варіації вибірових оцінок.

Для визначення оцінок дисперсії, стандартної та відносної похибок показників використовуються спеціалізовані загальноприйняті статистичні методи розрахунків.

При аналізі якості даних ОУЖД застосовується реплікаційний метод збалансованих повторних дублювань (balanced repeated replications – BRR), який реалізований у стандартному пакеті програм WesVarPC. BRR є досить складною розрахунковою процедурою. Вона передбачає, що повна вибірка складається з H шарів, у кожному з яких міститься дві територіальні одиниці. Кожна реплікація – «напіввибірка» будується шляхом випадкового відбору по одній територіальній одиниці з кожного шару.

Технологія розрахунку похибки вибірки для результатів ОУЖД орієнтована на використання стандартних пакетів SPSS, WesVarPC, EXCEL. Для розрахунку характеристик надійності основних показників доходів і витрат за результатами ОУЖД у файлі аналізу даних обстеження, що містить інформацію по домогосподарствах або по членах домогосподарств, формуються необхідні для розрахунків змінні:

- 1) Змінні характеристики основних груп доходів і витрат:
 - грошові доходи (cashinc);
 - грошові витрати (cashexp);
 - сукупні витрати (totalexpr);
 - загальні доходи (totalinc);
 - сукупні ресурси (resurs).
 - 2) Класифікаційні змінні, для яких розраховуються основні показники доходів і витрат:
 - тип населеного пункту (tp_ns_p)
 - код територій (регіонів) за КОАТУУ (cod_obl);
 - код економічного району (rn).
 - 3) Спеціальні змінні, що використовуються програмою при розрахунках:
 - порядковий номер запису (id) ;
 - статистична вага (w_q);
 - страти для України (strt_ua);
 - страти для регіонів (strt_rg);
 - страти для економічних районів (strt_rn);
 - код шару первинної одиниці вибірки (psu_ua).
- У діалогових вікнах програми WesVarPC задаються:
- набір необхідних для розрахунків змінних;
 - метод розрахунку;
 - змінні страт і шарів;
 - вихідні таблиці з характеристиками надійності.

Результати роботи програми зберігається у виводу WesVarPC (*.lst), який завантажується через середовище Excel для наступного аналізу та формування таблиць у заданих форматах:

- 1). У верхньому лівому куті листа міститься інформація про програму, дату розрахунку, місцезнаходження вхідної та вихідної інформації та загальні параметри для розрахунку.
- 2). Нижче послідовно виводяться таблиці з результатами розрахунків, які містять таку інформацію:
 - класифікаційні змінні, для яких побудовані комірки таблиць;
 - інформація про характер і тип оцінки (графи «STATISTIC» та «EST_TYPE»);
 - оцінка показника (графа «ESTIMATE»);
 - стандартна похибка оцінки (графа «STDERROR»);
 - коефіцієнт варіації (відносна стандартну похибку) оцінки (графа «CV»);
 - розмір групи, для якої розраховується оцінка показника (графа «N»);
 - оцінка величини дизайн-ефекту (графа «DEFF»).

Основні етапи роботи щодо розрахунку характеристик надійності основних показників доходів і витрат за даними ОУЖД:

1). У програмному середовищі SPSS готується робочий файл аналізу даних по домогосподарствах, який містить цільові показники, ключові змінні (код домогосподарства, загальнодержавний код ПТОВ), класифікаційні змінні (тип населеного пункту) та статистичні ваги.

2). У робочому файлі аналізу даних формуються спеціальні змінні, необхідні для розрахунків. Це здійснюється автоматизовано шляхом використання функції SPSS злиття робочого файла аналізу даних зі спеціальним масивом, який містить спеціальні змінні. Ключовою змінною для злиття масивів є загальнодержавний номер ПТОВ.

3). Робочий файл аналізу даних імпортується у формат програми WesVarPC (створюється файл із розширенням *.var) та подальші розрахунки проводяться у середовищі цієї програми.

- 4). Розраховуються реплікаційні ваги домогосподарств.
- 5). Здійснюється розрахунок оцінок показників і характеристик їх надійності.
- 6). Файл із результатами розрахунків відкривається в середовищі EXCEL, де проводиться формування остаточних таблиць для публікацій результатів ОУЖД.

3.2.6. Підготовка інформаційно-аналітичних матеріалів за результатами обстеження

Після проходження всіх етапів обробки на основі даних файлів аналізу даних формуються аналітичні таблиці для публікацій статистичних збірників і бюлетенів з результатами ОУЖД. За допомогою спеціально розробленої синтаксис-програми SPSS розраховуються абсолютні показники, згруповані за типами домогосподарств:

- за типом населеного пункту;
- за кількістю осіб у складі домогосподарства,
- за наявністю дітей у складі домогосподарства,
- за кількістю дітей у домогосподарствах з дітьми,
- за кількістю дорослих осіб у складі домогосподарства з дітьми,
- за наявністю у складі домогосподарства дітей, які не мають одного чи обох батьків,
- за кількістю осіб у домогосподарствах без дітей,
- за наявністю працюючих осіб,
- залежно від віку та статі особи, яка очолює домогосподарство (голови домогосподарства),
- за середньодушовими грошовими доходами у місяць,
- за середньодушовими загальними доходами у місяць,
- за децильними групами залежно від розміру середньодушового грошового доходу,
- за децильними групами залежно від розміру середньодушового загального доходу,
- за регіонами,
- за економічними районами.

Подальші розрахунки відносних показників (середніх тощо) виконуються в програмному середовищі Excel. При цьому для кожної аналітичної таблиці розроблені спеціальні арифметичні та логічні контролі.

1). Значення сумарних показників повинні дорівнювати сумі відповідних складових показників, тобто витрати на транспорт, наприклад, дорівнюють сумі витрат на послуги транспорту та на товари з цієї групи. У таблицях для публікацій, що містять дані щодо витрат, наводяться загальні показники (продовольчі товари, непродовольчі товари, послуги), які мають дорівнювати сумі відповідних складових по окремих групах КІСЦ. Наприклад, показник «послуги» повинен дорівнювати сумі показників по всіх групах послуг, наведених у таблиці. Такий тип контролю застосовується як для середніх показників, так і для структурних показників.

2). У публікаціях в окремі розділи виділено витрати домогосподарств на продукти харчування за окремими видами продуктів. Для контролю правильності розрахунку побудовано контроль між розділами (загальна сума витрат на продукти харчування в цьому окремому розділі повинна дорівнювати показнику витрат на продукти харчування в таблицях, наведених у розділі щодо сукупних витрат домогосподарств).

3). Має бути здійснена перевірка арифметичної та логічної відповідності розрахованих показників. Наприклад, значення середнього показника по Україні не може бути меншим за найменше значення показника для регіонів та більшим за найбільше серед регіонів значення показника. Такий контроль здійснюється для всіх груп показників витрат, доходів і ресурсів домогосподарств, а також для груп показників щодо характеристики споживання продуктів харчування. Приклад перевірки логічної відповідності: у домогосподарствах з дітьми зазвичай витрати на одяг та взуття, освіту, відпочинок і культуру, харчування поза домом більші, ніж серед домогосподарств без дітей. Або ж вартість спожитої продукції, отриманої з особистого підсобного господарства та від самозаготівель, серед домогосподарств сільської місцевості значно перевищує аналогічні показники для міських домогосподарств. Ці види контролю дозволяють перевірити правильність побудови таблиць і розрахунку показників у залежності від типу домогосподарства.

5) Також проводиться аналіз логічної узгодженості основних показників між регіонами. Наприклад, сукупні ресурси в розрахунку на одного члена домогосподарства Тернопільської або Волинської областей не можуть перевищувати аналогічний показник для м. Києва.

Аналітичні таблиці являються основою для підготовки публікацій результатів ОУЖД. Нижченаведені типи публікацій і періодичність їх видання.

I. Статистичні бюлетені:

1. «Витрати і ресурси домогосподарств України» – *щоквартально*.
2. «Витрати і ресурси домогосподарств регіону» (27 публікацій) – *раз на рік*.
3. «Витрати і ресурси домогосподарств економічного району» (8 публікацій) – *раз на рік*.

II. Статистичні збірники:

1. «Соціально-демографічні характеристики домогосподарств України» – *раз на рік*.
2. «Витрати і ресурси домогосподарств України за рік» – *раз на рік*.
3. «Самооцінка населенням стану здоров'я та рівня доступності окремих видів медичної допомоги» – *раз на рік*.
4. «Наявність у домогосподарствах товарів тривалого користування» – *раз на 2 роки*.
5. «Самооцінка домогосподарствами України рівня своїх доходів» – *раз на рік*.

III. Доповіді:

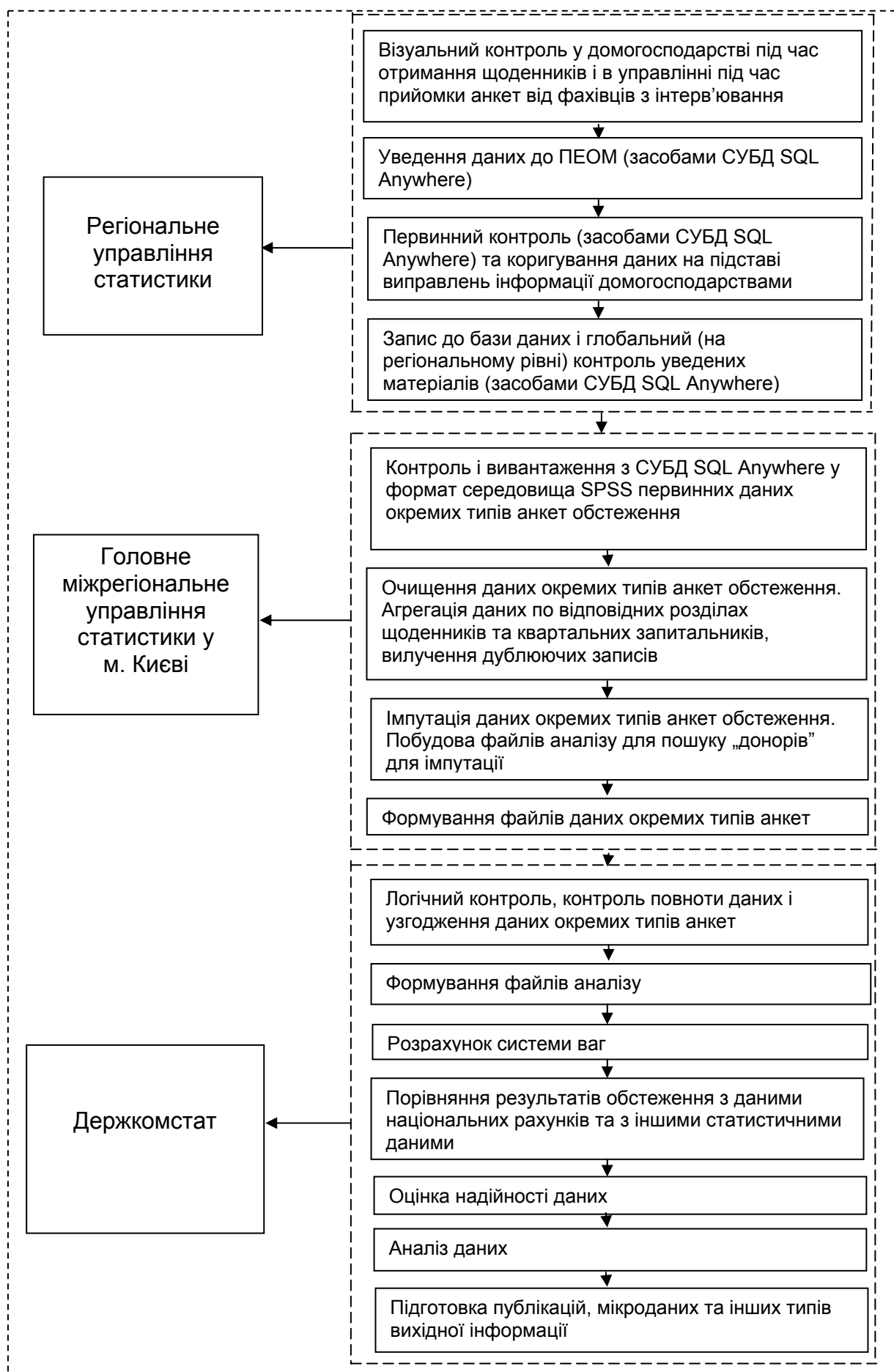
1. «Соціально-демографічні характеристики домогосподарств України» – *раз на рік*.
2. «Витрати і ресурси домогосподарств України» – *за I квартал, I півріччя та 9 місяців*.
3. «Соціально-економічне становище домогосподарств України» – *у цілому за рік*.
4. «Самооцінка населенням стану здоров'я та рівня доступності окремих видів медичної допомоги» – *раз на рік*.
5. «Наявність у домогосподарствах товарів тривалого користування» – *раз на 2 роки*.
6. «Самооцінка домогосподарствами доступності окремих видів товарів та послуг» – *раз на 2 роки*.
7. «Самооцінка домогосподарствами України рівня своїх доходів» – *раз на рік*.

До публікацій також додаються методологічні коментарі щодо програми обстеження, основних принципів побудови вибірки, поширення даних на генеральну сукупність, аналізу рівня участі домогосподарств в обстеженні, аналізу надійності оцінок основних показників, пояснення щодо основних термінів і понять.

Окрім випуску результатів обстеження у вигляді публікацій, оприлюднення основних результатів відбувається на офіційному сайті Держкомстату у відповідному розділі: www.ukrstat.gov.ua, де, зокрема, розміщуються і звіти з якості результатів ОУЖД. На запити користувачів надаються мікрофайли даних із результатами ОУЖД і описи цих файлів.

Для потреб розрахунків макроекономічних показників і показників інших галузей статистики за результатами ОУЖД готуються додаткові аналітичні таблиці (див. додаток б).

Схема етапів робіт з обробки даних обстеження умов життя домогосподарств



Строки виконання етапів робіт з обробки даних ОУЖД Головним міжрегіональним управлінням статистики у м. Києві у 2011 році

	Контрольна картка + Запитальник основного інтерв'ю	Щоденник поточних витрат домогосподарств	Квартальне інтерв'ю + Анкета №1 «витрати на будівництво та ремонт»	Анкета №2 «Стан здоров'я» + «Доступ до товарів та послуг / Наявність товарів тривалого користування»	Анкета №3 «Самооцінка домогосподарствами своїх доходів»
Отримання файлів від регіональних управлінь	10.01	15.04 – I кв., 15.07 – II кв., 14.10 – III кв., 21.01 – IV кв.	20.05 – I кв., 16.08 – II кв., 16.11 – III кв., 23.02 – IV кв.	16.11 – III кв.	23.02 – IV кв.
Контроль і вивантаження даних у формат SPSS	17.01	22.04 – I кв., 22.07 – II кв., 21.10 – III кв., 04.02 – IV кв.	26.05 – I кв., 19.08 – II кв., 18.11 – III кв., 04.03 – IV кв.	24.11 – III кв.	04.03 – IV кв.
Очищення даних	31.03	13.05 – I кв., 12.08 – II кв., 11.11 – III кв., 11.03 – IV кв.	10.06 – I кв., 15.09 – II кв., 15.12 – III кв., 01.04 – IV кв.	29.12 – III кв.	14.04 – IV кв.
Імпутація відсутніх даних	07.04	23.06 – I кв., 22.09 – II кв., 22.12 – III кв., 08.04 – IV кв.	24.06 – I кв., 23.09 – II кв., 29.12 – III кв., 15.04 – IV кв.		
Формування файлів для передачі до Держкомстату	08.04 – по особах; 14.04 – по д/г	23.06 – I кв., 22.09 – II кв., 22.12 – III кв., 08.04 – IV кв.	24.06 – I кв., 23.09 – II кв., 29.12 – III кв., 15.04 – IV кв.		
Формування файла контрольної картки домогосподарств з урахуванням змін у складі домогосподарства	05.06 – I кв., 04.09 – II кв., 04.12 – III кв., 27.03 – IV кв.				

Схема етапів робіт з контролю та очищення даних статистичних формулярів



Застосування процедур імпутації при обробці даних ОУЖД

Тип статистичного формуляра	Наявність імпутації	Рівень імпутації	Метод імпутації
Запитальник основного інтерв'ю	Є	Усі змінні, які мають НЗВ	Найближчого сусіда
Щоденник поточних витрат	Є	Пропущені дні та тижні заповнення щоденника	Найближчого сусіда
Запитальник квартального інтерв'ю	Є	Розділ III. Житлово-комунальні витрати: – домогосподарства, що проживають у квартирах, будинках; – домогосподарства, що проживають у гуртожитках	Найближчого сусіда, середнього значення
Анкета №1 – Будівництво та ремонт житла	Немає	<u>Не спостерігається НЗВ</u> , у рідкісних випадках пропущені значення імпутуються методом найближчого сусіда або середнього значення	–
Анкета №2, ч.1 – Стан здоров'я членів домогосподарства	Немає	<u>Не спостерігається НЗВ</u> , у рідкісних випадках пропущені значення імпутуються методом найближчого сусіда	–
Анкета №2, ч.2 – Наявність товарів тривалого користування / Доступ до окремих товарів та послуг	Немає	<u>НЗВ не імпутуються</u> , у публікаціях результатів указується рівень НЗВ по окремих змінних	–
Анкета №3 – Самооцінка рівня своїх доходів	Немає	<u>НЗВ не імпутуються</u> , у публікаціях результатів указується рівень НЗВ по окремих змінних	–

Строки виконання етапів робіт з обробки даних ОУЖД у департаменті обстежень домогосподарств Держкомстату у 2011 році

	Контрольна картка + Основне інтерв'ю	Щоденник поточних витрат	Квартальне інтерв'ю + витрати на будівництво та ремонт	Самооцінка домогосподарствами своїх доходів	Стан здоров'я + Доступ до товарів та послуг / Наявність товарів тривалого користування
Розрахунок системи ваг	05.05	07.07 – I кв.; 29.09 – II кв.; 05.01 – III кв.; 21.04 – IV кв.			
Формування файлів аналізу	12.05	14.07 – I кв.; 07.10 – II кв.; 21.10 – I півріччя; 14.01 – III кв.; 28.01 – 9 міс.; 29.04 – IV кв.; 26.05 – рік		26.05	11.02
Підготовка аналітичних таблиць і публікацій	19.05	18.08 – I кв.; 27.10 – II кв.; 28.01 – III кв.; 19.05 – IV кв.		15.06	23.02

Результати ОУЖД, які надаються галузевим департаментам Держкомстату

Департамент	Інформація	Розріз	Періодичність
<u>Макроекономічної статистики</u>	Структура споживчих грошових витрат домогосподарств на придбання продовольчих, непродовольчих товарів та оплату послуг за каналами реалізації	Україна, регіони, економічні райони, Центральний екон. район без м. Києва	Річна
	Витрати домогосподарств на оплату послуг, наданих громадянами	Україна, економічні райони	Квартальна, річна
	Доходи домогосподарств від продажу нерухомості	Україна, економічні райони	Квартальна, річна
	Доходи домогосподарств від продажу особистого і домашнього майна	Україна, економічні райони	Квартальна, річна
	Розподіл осіб, які входять до складу домогосподарств, за статусом їх зайнятості в розрізі підсекторів сектору домогосподарств	Україна	Річна
<u>Статистики цін</u>	Збір лісових продуктів, рибальство, мисливство	Україна	Квартальна, річна, Річна
	Структура споживчих грошових витрат домогосподарств на придбання продовольчих, непродовольчих товарів та оплату послуг за каналами реалізації	Україна, регіони, економічні райони, Центральний екон. район без м. Києва	Річна
	Структура споживчих грошових витрат домогосподарств	Україна, економічні райони, м. Київ, Центральний екон. район без м. Києва	Річна
<u>Статистики торгівлі</u>	Споживчі грошові витрати домогосподарств (у розрахунку на 100 домогосподарств)	Україна, економічні райони, м. Київ, Центральний екон. район без м. Києва	Річна
	Структура споживчих грошових витрат домогосподарств на придбання продовольчих, непродовольчих товарів та оплату послуг за каналами реалізації	Україна Україна регіони, економічні райони, Центральний екон. район без м. Києва	Квартальна, Річна
<u>Статистики послуг</u>	Купівля продовольчих товарів (у перерахунку в основний продукт)	Україна	Річна
	Витрати населення на одержання послуг на підприємствах та у громадян	Україна	Річна

Департамент	Інформація	Розріз	Періоди- чність
<u>Сільського господарства та навколишнього середовища</u>	Витрати домогосподарств, пов'язані з сільськогосподарською діяльністю Збір лісових продуктів, рибальство, мисливство	Україна, тип місцевості, регіони Україна, регіони за типом місцевості	Квартальна Річна
	Наявність та рух поголів'я худоби та птиці в домогосподарствах	Україна, тип місцевості, регіони Україна, регіони за типом місцевості	Квартальна Річна
	Використання продуктів тваринництва	Україна, тип місцевості, регіони Україна, регіони за типом місцевості	Квартальна Річна
	Надходження від продажу сільськогосподарських продуктів	Україна, тип місцевості, регіони Україна, регіони за типом місцевості	Квартальна Річна
	Вирощування та використання сільськогосподарських культур	Україна, тип місцевості, регіони Україна, регіони за типом місцевості	Квартальна Річна