

Державна служба статистики України

ЗАТВЕРДЖЕНО

**Наказ Державної служби
статистики України**

07.11.2017 № 286

**МЕТОДИКА
ЗАСТОСУВАННЯ ВИБІРКОВОГО МЕТОДУ
В ДЕРЖАВНОМУ СТАТИСТИЧНОМУ СПОСТЕРЕЖЕННІ
"ВИКОРИСТАННЯ ТА ЗАПАСИ ПАЛИВА"**

Київ – 2017

Державна служба статистики України

Відповідальний за випуск – директор департаменту статистики торгівлі Фризоренко А. О.

Методика застосування вибіркового методу в державному статистичному спостереженні "Використання та запаси палива" (далі – Методика) містить опис методу формування вибіркової сукупності для обстеження одиниць державного статистичного спостереження щодо використання та запасів палива з місячною періодичністю. У Методиці описано джерела інформації, дизайн вибірки та процедури формування основи вибірки, виявлення нетипових одиниць, обчислення обсягу вибірки в цілому та за стратами, процедуру відбору одиниць для безпосереднього обстеження.

Методику підготували фахівці департаменту статистики торгівлі за участю фахівців департаменту статистичної інфраструктури Держстату.

Методика схвалена Комісією з питань удосконалення методології та звітної документації Державної служби статистики України (протокол від 23 грудня 2016 року № 16).

Державна служба статистики України

- адреса: вул. Шота Руставелі, 3, Київ, 01601
- телефон: (044) 287-24-22
- факс: (044) 235-01-53
- електронна пошта: office@ukrstat.gov.ua
- веб-сайт: www.ukrstat.gov.ua

Зміст

	Стор.
Передмова.....	4
Перелік умовних позначень і скорочень.....	5
I. Загальні положення	6
II. Визначення основних понять і термінів	6
III. Формування генеральної сукупності та основи вибірки	7
IV. Стратифікація основи вибірки.....	8
V. Виділення нетипових одиниць.....	8
VI. Обсяг вибірки.....	9
VII. Відбір одиниць вибіркової сукупності	11
VIII. Обчислення системи статистичних ваг	12
IX. Отримання оцінок показників за результатами вибіркового спостереження	13
X. Перевірка надійності показників вибіркового спостереження	14
Додатки.....	
Додаток 1. Загальна схема плану (дизайну) вибірки для державного статистичного спостереження "Використання та запаси палива".....	16
Список використаних джерел.....	17

Передмова

Методика застосування вибіркового методу в державному статистичному спостереженні "Використання та запаси палива" за формою № 4-мп (місячна) (далі – Методика) підготовлена на виконання Стратегії розвитку державної статистики на період до 2017 року.

Методика містить опис методу формування вибіркової сукупності для обстеження одиниць щодо використання та запасів палива з місячною періодичністю. Наведено опис джерел інформації, методику формування основи вибірки, дизайну вибірки, виявлення нетипових одиниць, обчислення обсягу вибірки в цілому та за втратами, розрахунок і коригування статистичних ваг одиниць спостереження, відбору одиниць для безпосереднього обстеження.

Методика призначена для використання у практичній діяльності органів державної статистики.

Перелік умовних позначень і скорочень

Перелік скорочень

КОАТУУ – Класифікатор об'єктів адміністративно-територіального устрою України;

PCO – Реєстр статистичних одиниць.

Перелік позначень

- N – кількість одиниць у генеральній сукупності;
- n_i – одиниця вибірки;
- n – загальний обсяг вибірки;
- $N_j^{IV,V}$ – кількість одиниць у стратах IV, V для j -го регіону;
- $n_j^{IV,V'}$ – обсяг вибірки зі страт IV, V для j -го регіону;
- y_{hji} – значення оцінюваного показника для i -ї одиниці у j -му регіоні;
- \bar{y}_{hj} – середнє значення оцінюваного показника для j -го регіону;
- m – крок відбору;
- w_i – базова вага i - i одиниці;
- w_i' – кінцева вага i - i одиниці;
- $n_j^{\text{невід.}}$ – кількість одиниць із причинами невідповідей;
- \hat{Y}_j – оцінка сумарного значення оцінюваного показника для j -го регіону;
- $SE(\hat{Y}_j)$ – оцінка сумарного значення стандартної похибки оцінюваного показника для j -го регіону у стратах IV, V;
- $SE(\bar{y}_j)$ – середнє значення стандартної похибки оцінюваного показника для j -го регіону у стратах IV, V;
- CV – коефіцієнт варіації;
- k – коефіцієнт невідповідей;
- $\sigma_j^{IV,V}$ – середньоквадратичне відхилення оцінюваного показника у стратах IV, V для j -го регіону.

Оцінюваним показником вибіркового спостереження є використання палива за певними видами, зазначеними у пункті 3 розділу III.

I. Загальні положення

1. Державне статистичне спостереження "Використання та запаси палива" за формою № 4-мп (місячна) (далі – спостереження) здійснюється на постійній основі з місячною періодичністю і базується на європейських стандартах, зокрема Регламенті (ЄС) № 1099/2008 Європейського парламенту та ради від 22 жовтня 2008 року про статистику в галузі енергетики.

Метою спостереження є отримання даних щодо обсягів використання та запасів палива на вибірковій основі.

2. Формування вибіркової сукупності здійснюється на державному рівні на основі процедури стратифікованого систематичного відбору одиниць з урахуванням обсягів використання палива.

3. Ця методика розкриває методи побудови вибіркової сукупності державного статистичного спостереження для отримання результатів щодо використання та запасів палива та оцінки їх надійності.

4. Реалізація процедур здійснюється з використанням пакета обробки статистичних даних SPSS IBM та Microsoft Excel.

II. Визначення основних понять і термінів

1. Для цілей цієї Методики терміни вживаються у значеннях, визначених Глосарієм до плану статистичного спостереження [1], термінологічним словником щодо вибірових обстежень[2], а також у значеннях:

1) базова вага – кількість одиниць генеральної сукупності, яку за планом вибірки повинна представляти відібрана у вибірці одиниця;

3) домен – підсукупність одиниць генеральної сукупності, для якої необхідно оцінити деякі статистичні показники. В обстеженні за місячним використання палива доменами є сукупності підприємств за видами палива, регіонами;

4) кінцева вага – кількість одиниць генеральної сукупності, яку репрезентує одиниця, що була обстежена;

5) невідповідь – відсутність інформації за деякими одиницями спостереження, що потрапили до вибірки;

б) перезважування – процедури, пов'язані з перерахунком базових ваг на коригуючі коефіцієнти, у результаті яких отримуються кінцеві ваги.

III. Формування генеральної сукупності та основи вибірки

1. Одиницею спостереження є місцева одиниця за місцем здійснення діяльності, яка використовує та/або має у запасах паливо.

2. Сукупність місцевих одиниць для обстеження використання та запасів палива формується за даними реєстру статистичних одиниць (далі – РСО) з урахуванням додаткових критеріїв щодо показників паливно-енергетичних ресурсів.

Виділення великих та малих підприємств, для яких традиційно використовуються різні методи обстеження, не є характерними для цього спостереження, оскільки малі підприємства можуть бути значними споживачами палива. Тому спостереженням охоплюються одиниці незалежно від розміру.

3. Програмою спостереження передбачено щомісячне обстеження 8-ми видів палива (вугілля кам'яне, вугілля буре, нафта сира, включаючи газовий конденсат, газ природний, бензин моторний, газойлі (паливо дизельне), мазути паливні важкі та пропан й бутан скраплені) за показниками щодо обсягів їх використання та запасів. Між собою ці показники за видами палива не корелюються, тому кожний вид палива необхідно розглядати окремо.

Ураховуючи, що використання таких видів палива, як вугілля буре, нафта сира, включаючи газовий конденсат, та мазути паливні важкі здійснюється незначним колом одиниць, то вони обстежуються на суцільній основі. Тому для вибіркового методу прийнято рішення обмежитися 5-ма угрупованнями видів палива: вугілля кам'яне, газ природний, бензин моторний, паливо дизельне, пропан й бутан скраплені.

3. Формування основи вибірки відбувається у три етапи:

1) формування сукупності одиниць, які використовують 8 видів палива (вугілля буре, нафта сира, включаючи газовий конденсат, та мазути паливні важкі, вугілля кам'яне, газ природний, бензин моторний, паливо дизельне, пропан й бутан скраплені);

2) формування сукупності одиниць, які мають у запасах паливо і представляють 80% від загальних річних обсягів по кожному виду палива;

3) об'єднання сукупностей одиниць, зазначених у пунктах 3.1 та 3.2 цього розділу, у єдину основу вибірки.

4. Відбір одиниць відбувається за ідентифікаційним кодом місцевої одиниці, при цьому дублі місцевих одиниць вилучаються.

IV. Стратифікація основи вибірки

1. Процедура стратифікації основи вибірки полягає у попередньому розподіленні основи вибірки на регіони та на більш однорідні групи (страти) з подальшим відбором певної кількості одиниць для безпосереднього обстеження у кожному з регіонів.

Стратифікації піддаються одиниці всієї сукупності за 8-ми видами палива і за характерними ознаками виділяються такі страти:

1) страта I (N^I) – сукупність одиниць, які використовують вугілля буре, нафту сиру, включаючи газовий конденсат та мазути паливні важкі, які обстежуються суцільно (відбирається 100 % одиниць);

2) страта II (N^{II}) – сукупність одиниць, які мають лише у запасах будь-який із видів палива спостереження (відбирається 100%);

3) страта III (N^{III}) – сукупність нетипових одиниць, які мають значні обсяги використання 5-ти видів палива – вугілля кам'яного, газу природного, бензину моторного, палива дизельного, пропану й бутану скраплених (відбирається 100%);

4) страта IV (N^{IV}) – сукупність одиниць, які не мають даних щодо використання та запасів палива, так звані "нульові одиниці" (відбирається 10%);

5) страта V (N^V) – сукупність одиниць, які мають дані щодо використання палива (вугілля кам'яного, газу природного, бензину моторного, палива дизельного, пропану й бутану скраплених), за обсяги меншими порівняно з тими, що відібрані до старти III і визначені як "нетипові" (відбирається 20 %);

б) загальний обсяг вибірки розподіляється за регіонами. Отримані результати обсягу вибірки округлюють до цілих чисел.

V. Виділення нетипових одиниць

1. Нетипові одиниці виділяються по кожному регіону окремо.

Для виявлення нетипових одиниць за кожним з 5-ти видів палива відбувається в порядку зростання обсягів використання певного виду палива з подальшою оцінкою їхніх значень математичними методами.

2. Нетиповими вважають підприємства, для яких значення використання певного виду палива потрапляє за межі інтервалу:

$$|y_{nji} - \bar{y}_{nj}| \geq 3\sigma_{nj} \quad (1)$$

де Y_{hji} – значення обсягів використання для i -ї одиниці виду палива (h) для j -го регіону;
 \bar{Y}_{hj} – середнє значення обсягів використання певного виду палива (h) для j -го регіону;
 σ_{hj} – середньоквадратичне відхилення обсягу використання певного виду палива для j -го регіону.

VI. Обсяг вибірки

1. Обчислення обсягу вибірки відбувається у два етапи:

- 1) обчислення загального обсягу вибірки;
- 2) обчислення обсягу вибірки за стратами та його розподіл за регіонами.

За результатами досліджень, із урахуванням оцінок точності параметрів, за показником щодо використання палива для національного та регіонального рівнів загальний обсяг вибіркової сукупності є достатнім на рівні 20% генеральної сукупності.

2. Загальний обсяг вибірки визначається:

$$n = N \cdot 0,2 \quad (2)$$

3. Загальний обсяг вибірки складають:

- одиниці, визначені у страті I (відбираються 100 %);
- одиниці, визначені у страті II (відбираються 100%);
- одиниці, визначені у страті III за обсягами використання вугілля кам'яного, газу природного, бензину моторного, газойлів (палива дизельного), які є "нетиповими" (відбираються 100%);
- одиниці, визначені у страті IV з нульовими показниками (відбираються 10%);
- одиниці, визначені у страті V, для яких обсяг використання бензину моторного, вугілля кам'яного, газу природного, газойлів (палива дизельного), пропану й бутану скраплених є "типовим" (відбираються в обсязі, визначеному за формулою (6)).

Загальна схема плану (дизайну) вибірки наведена в додатку 3.

4. Для визначення вибіркової сукупності одиниць страти III (N^{III}) по 5-ти видах палива по кожному регіону визначають "нетипові" одиниці, використовуючи математичний метод трьох сигм (3σ).

"Нетипові" одиниці групуються по кожному регіону і включаються до вибіркової сукупності зі 100% відбором:

$$n^{III} = N^{III} = \sum_1^5 n_{hj}^{N^{III}}, \quad (3)$$

де n^{III} - кількість "нетипових" одиниць у страті III (N^{III});
 h - вид палива, що підлягає вибірковому обстеженню.

5. Потім визначається обсяг одиниць вибірки для страти IV в цілому для України :

$$n^{IV} = N^{IV} \cdot 0,1 \quad (4)$$

де N^{IV} – загальна кількість одиниць у страті IV.

6. Обсяг одиниць вибірки для страти V в цілому по Україні визначається наступним чином:

$$n^V = N \cdot 0,2 - n^I - n^{II} - n^{III} - n^{IV} \quad (5)$$

Після чого отриманий обсяг вибірки V - ої страти з урахуванням невідповідей, відсоток яких встановлено на рівні 2% (за даними обстеження за попередні роки), коригується таким чином:

$$n^{V'} = n^V \cdot 0,02 \quad (6)$$

6. Обчислюється обсяг вибірки страт IV та V по кожному регіону з використанням формули оптимального розміщення Неймана:

$$n_j^{IV,V'} = n^{IV,V'} \cdot \frac{\sigma_j^{IV,V'} \cdot N_j^{IV,V'}}{\sum_{j=1}^{25} \sigma_j^{IV,V'} \cdot N_j^{IV,V'}} \quad (7)$$

де $n_j^{IV,V'}$ – обсяг вибірки зі страт IV, V для j -го регіону з урахуванням невідповідей;

$N_j^{IV,V'}$ – кількість одиниць у стратах IV, V для j -го регіону;

$\sigma_j^{IV,V'}$ – середньоквадратичне відхилення використання палива у стратах IV, V для j -го регіону.

Для обчислення визначених у п. 5 та п. 6 обсягів вибірки за регіонами розраховується сума обсягів вибірки для страт IV, V та розповсюджується за J -м регіоном.

7. Із отриманої вибіркової сукупності видаляються дублі одиниць. Розрахунковий обсяг вибірки округлюють до цілого числа.

VII. Відбір одиниць вибіркової сукупності

1. Відбір одиниць здійснюється поетапно з ймовірністю відбору:

$$p = \frac{n}{N} \quad (8)$$

де n – кількість одиниць вибірки;

N – кількість одиниць генеральної сукупності.

2. Спочатку для кожного регіону відбираються одиниці, які задовольняють умовам, визначеним для страт n^I, n^{II}, n^{III} з коефіцієнтом ймовірного відбору, що дорівнює 1 ($p=1$).

3. Відбір одиниць страт IV, V здійснюється з використанням простого випадкового відбору.

Для цього визначається випадкове число, щоб визначити першу одиницю і крок відбору (m). Крок відбору визначається шляхом ділення обсягу сукупності на обсяг вибірки кожної страти:

$$m = \frac{N^{IV,V}}{n^{IV,V}} \quad (9)$$

Для розрахунку першої одиниці відбору спочатку розраховують змінну, помноживши крок відбору на випадкове число. Випадкове число (rnd) визначають за допомогою генератора випадкових чисел:

$$n_1 = m \cdot rnd \quad (10)$$

де rnd – випадкове число, визначене для страти, яке приймає значення від 0 до 1.

4. Перша одиниця відбору визначається відповідно до змінної. Якщо ця змінна більше або дорівнює 1, то вона буде вказувати першу одиницю відбору, а якщо при множенні кроку відбору на мале випадкове число розрахована змінна менша за 1, то в такому випадку для визначення номера першої одиниці до отриманої розрахункової змінної ще раз додається крок відбору.

5. Наступні одиниці спостереження визначають шляхом систематичного додавання кроку відбору до попередньо розрахованого номера одиниці у страті, округляючи результат до цілого числа:

$$n_{i+m} = n_1^{IV,V} + m^{IV,V} \quad (11)$$

6. Основа вибірки ранжується в порядку значень чисел від більшого до меншого.

VIII. Обчислення системи статистичних ваг

1. Розрахунок системи статистичних ваг здійснюється для страт IV, V у межах кожної страти регіону і складається з таких етапів:

- 1) розрахунок базової ваги;
- 2) коригування ваг для врахування ступеня участі відібраних одиниць у спостереженні;
- 3) контроль якості системи статистичних ваг.

2. Для кожної одиниці вибіркової сукупності необхідно обчислити її базову вагу. Базова вага дорівнює одиниці для одиниць вибіркової сукупності, які обстежуються суцільно (страти I, II, III).

Для всіх інших одиниць базові ваги визначаються за формулою:

$$w_{ji}^{IV,V} = \frac{N_j^{IV,V}}{n_j^{IV,V}} \quad (12)$$

де $N_j^{IV,V}$ – кількість одиниць у страті IV або V для j -го регіону;
 $n_j^{IV,V}$ – обсяг вибірки у страті IV або V для j -го регіону.

3. Для точної ідентифікації одиниць, які не відповіли, проводять деталізацію причин неподання звітів. Коди причин неподання звітів проставляють у відповідності до Довідника причин неподання форм державних статистичних спостережень.

Коригування базових ваг з урахуванням коефіцієнта невідповідей відбувається за формулою:

$$w'_i = w_i \cdot k \quad (13)$$

де w_i – базова вага i -ї одиниці;
 w'_i – кінцева вага i -ї одиниці;
 k – коефіцієнт невідповідей, який визначають для кожного регіону (j) за формулою:

$$k_j = \frac{n'_j}{n'_j - n_j^{\text{невід.}}} \quad (14)$$

де $n_j^{\text{невід.}}$ – кількість одиниць із причинами невідповідей (3 "Одиниця не відповідає критеріям одиниці спостереження", 6 "Одиниця призупинила діяльність" та 9 "Одиниця має сезонну діяльність") для j -го регіону.

4. З метою запобігання надмірного завищення ваг додатково аналізуються отримані відповіді по усіх одиницях вибірки.

5. Для оцінки якості статистичних ваг використовують формулу:

$$\sum_{i=1}^{n'_j} w'_{ij} = N_j; \quad N_j = 1 \quad (15)$$

ІХ. Отримання оцінок показників за результатами вибіркового обстеження

1. Отримання оцінок показників вибіркового обстеження за певним видом палива здійснюється шляхом зважування індивідуальних значень показників відібраних одиниць на їхні кінцеві ваги з використанням методу Горвіца – Томпсона за формулою:

$$\hat{Y}_j = \sum_{i=1}^{n'_j} w'_{ijh} \cdot y_{ijh} \quad (16)$$

де \hat{Y}_j – оцінка сумарного обсягу використання певного виду палива для j -го регіону у стратах IV або V;

y_{ijh} – значення показника використання певного виду палива для i -ї одиниці j -го регіону у стратах IV або V;

w'_{ijh} – кінцева вага i -ї одиниці j -го регіону у стратах IV або V.

2. Оцінки показників для державного рівня по кожному виду палива отримують як сумарне значення оцінок відповідних показників по регіонах.

Х. Перевірка надійності показників вибіркового обстеження

1. Для визначення якості результатів вибіркового обстеження оцінювання показників щодо обсягів використання певного виду палива (h) в межах кожного регіону визначають стандартні похибки за формулою:

$$SE(\hat{Y}_j) = N_j \cdot SE(\bar{y}_j) \quad (17)$$

де $SE(\hat{Y}_j)$ – оцінка сумарного значення стандартної похибки використання певного виду палива j -го регіону у стратах IV, V;
 $SE(\bar{y}_j)$ – середнє значення стандартної похибки використання певного виду палива j -го регіону у стратах IV, V;
 N_j – обсяг вибірки j -го регіону.

$$SE(\bar{y}_j) = \sqrt{\sum_{h=1}^5 \left(\left(\frac{N_h}{N_j} \right)^2 \cdot SE^2(\bar{y}_h) \right)}. \quad (18)$$

де h – вид палива у стратах IV, V ("нульові", "типові" одиниці)

$$SE(\bar{y}_h) = \frac{1}{\sqrt{n'_{jh}}} \cdot \sqrt{\frac{\sum_{i=1}^{n'_i} (y_{jhi} - \bar{y}_{jh})^2}{n'_{jh} - 1}} \cdot \sqrt{1 - \frac{n'_{jh}}{N_{jh}}} \quad (19)$$

2. Коефіцієнт варіації для j -го регіону визначається за формулою:

$$CV(\hat{Y}_j) = \frac{SE(\hat{Y}_j)}{\hat{Y}_j} \cdot 100\%, \quad (20)$$

де \hat{Y}_j – оцінка сумарного значення обсягів використання певного виду палива для j -го регіону.

3. Коефіцієнт варіації для України визначається за наступною формулою:

$$CV(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} \cdot 100\% \quad (21)$$

де \hat{Y} – оцінка сумарного значення обсягів використання певного виду палива по Україні.

4. Для визначення меж довірчих інтервалів для 95% рівня довірчої ймовірності розраховується гранична похибка вибірки. Наприклад, для оцінки середнього значення показника на регіональному рівні \bar{y}_j межі встановлюються за такою формулою:

$$ME(\bar{y}_j) = \pm 1,96 \cdot SE(\bar{y}_j) \quad (22)$$

5. Для визначення якості результатів вибіркового обстеження за показниками щодо використання певного виду палива по Україні розраховують стандартну похибку за формулою:

$$SE(\hat{Y}) = N \cdot SE(\bar{y}) \quad (23)$$

де $SE(\hat{Y})$ – оцінка сумарного значення стандартної похибки;
 $SE(\bar{y})$ – середнє значення стандартної похибки;
 N – загальний обсяг вибірки.

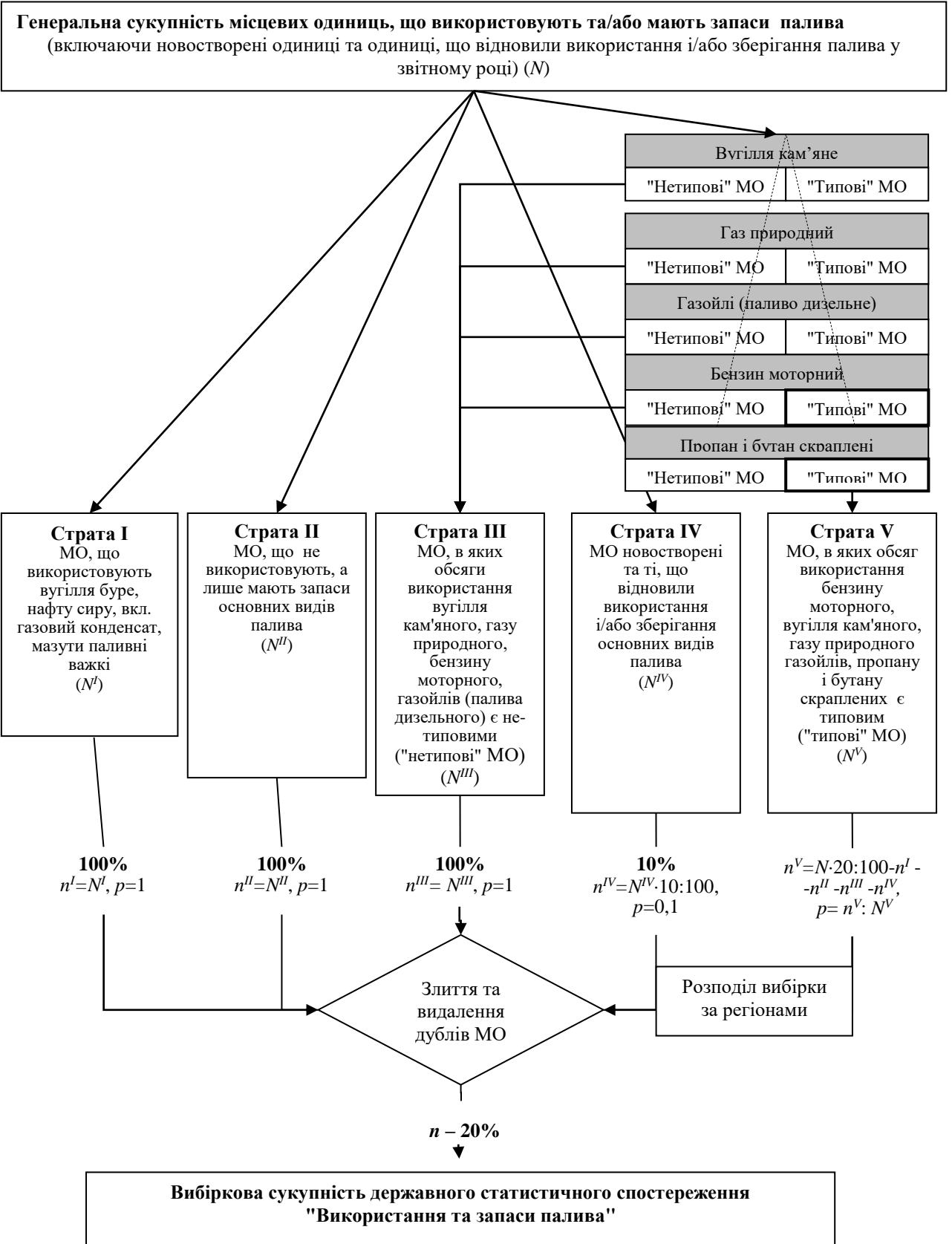
Середнє значення стандартної похибки певного виду палива по Україні визначають за формулою:

$$SE(\bar{y}) = \sqrt{\sum_{j=1}^{25} \left(\left(\frac{N_j}{N} \right)^2 \cdot SE^2(\bar{y}_j) \right)} \quad (24)$$

Директор департаменту
 статистики торгівлі

А. О. Фризоренко

Загальна схема плану (дизайну) вибірки для державного статистичного спостереження "Використання та запаси палива"



Список використаних джерел

1. Глосарій до плану статистичного спостереження, затверджений наказом Держкомстату від 29.12.2009 № 498 [Електронний ресурс]. – Режим доступу: <http://www.ukrstat.gov.ua>
2. Вибіркове спостереження: термінол. словн. / Укл. : О. О. Васечко, О. І. Черняк, Є. М. Жуйкова, О. В. Гончар та ін. – К. : Держкомстат України, 2004. – 140 с.
3. Класифікатор об'єктів адміністративно-територіального устрою України : у 3 т. : ДК 014:97. – [Чинний від 1997–10–31]. – К. : Держстандарт, 1997. – Т. 1–3. – 1196 с. (Державний класифікатор України).
4. Методи вибірових обстежень : [навч. посіб.] / Пархоменко В. М. – К. : [Б. в.], 2001. – 148 с.
5. Методика розрахунку системи статистичних ваг для поширення результатів державного вибіркового обстеження умов життя домогосподарств на генеральну сукупність; затвердж. Наказом Державного комітету статистики України від 14.11.2006 р. № 520 [Електронний ресурс]. – Режим доступу: <http://zakon.nau.ua/doc/?uid=1157.464.0>
6. Організація вибірових обстежень : [навч. посіб.] / Єріна А. М. – К. : КНЕУ, 2004. – 127 с.
7. Положення про Реєстр статистичних одиниць та формування основ вибірки сукупностей одиниць статистичних спостережень за діяльністю підприємств / Державний комітет статистики України – 2010.