

Державна служба статистики України

ЗАТВЕРДЖЕНО

Наказ Державної служби
статистики

07 лютого 2025 року № 24

**МЕТОДИКА
ФОРМУВАННЯ ВИБІРКОВОЇ СУКУПНОСТІ ОДИНИЦЬ
ДЕРЖАВНОГО СТАТИСТИЧНОГО СПОСТЕРЕЖЕННЯ
"ВИКОРИСТАННЯ ТА ЗАПАСИ ПАЛИВА" ТА ПОШИРЕННЯ
РЕЗУЛЬТАТІВ ВИБІРКОВОГО СПОСТЕРЕЖЕННЯ НА
ГЕНЕРАЛЬНУ СУКУПНІСТЬ**

Державна служба статистики України

Відповідальний за підготовку – директор департаменту статистики зовнішньоекономічної діяльності та енергетики апарату Держстату Валентина Кругляк.

Методика формування вибіркової сукупності одиниць державного статистичного спостереження "Використання та запаси палива" та поширення результатів вибіркового спостереження на генеральну сукупність (далі – Методика) містить опис процедур (етапів) формування сукупностей одиниць статистичного спостереження, що вивчаються, які використовують та/або мають у запасах паливо.

Методика призначена для використання працівниками органів державної статистики, а також може бути корисна для інших зацікавлених користувачів статистичної інформації.

Методика є складовою частиною Методологічних положень державного статистичного спостереження "Використання та запаси палива".

Методику підготували працівники відділу статистики енергетики департаменту статистики зовнішньоекономічної діяльності та енергетики апарату Держстату.

Методика схвалена Комісією з питань удосконалення методології та звітно-статистичної документації Держстату (протокол від 13 грудня 2024 року № КПУМ/24-24).

Держстат
вул. Шота Руставелі, 3, Київ, 01601
<http://www.ukrstat.gov.ua>
телефон: (044) 284-31-32, 289-76-40, 284-95-48
адреса електронної пошти: office@ukrstat.gov.ua

Зміст

	Стор.
Скорочення та умовні позначення.....	4
I. Вступ.....	5
II. Джерела інформації.....	6
III. Генеральна сукупність (основа вибірки) ДСС.....	6
IV. Дизайн вибірки та методичні підходи до її формування.....	7
V. Методичні підходи до поширення результатів обстеження на генеральну сукупність.....	10
VI. Методи компенсації відсутніх даних (імпутації).....	11
VII. Оцінювання показників за результатами ДСС.....	12
VIII. Визначення характеристик надійності оцінювання показників за результатами обстеження показників вибіркового обстеження.....	13
Додатки:	
Додаток 1.....	14
Додаток 2.	15
Додаток 3.	18
Список використаних джерел.....	38

Скорочення та умовні позначення

ДСС	– державне статистичне спостереження "Використання та запаси палива" за формою № 4-мтп (місячна) "Звіт про використання та запаси палива";
ДСС "Реєстр респондентів"	– державне статистичне спостереження "Реєстр респондентів статистичних спостережень";
Держстат	– Державна служба статистики України;
ЄС	– Європейський Союз;
КАТОТТГ	– Кодифікатор адміністративно-територіальних одиниць та територій територіальних громад;
КВЕД	– Класифікатор видів економічної діяльності
KICE	– Класифікація інституційних секторів економіки;
КОПФГ	– Класифікація організаційно-правових форм господарювання;
KICE	– Класифікація інституційних секторів економіки;
НДР	– науково-дослідна робота;
НАСОА	– Національна академія статистики, обліку та аудиту;
НКРЕКП	– Національна комісія, що здійснює державне регулювання у сферах енергетики та комунальних послуг;
НПП	– Номенклатура продукції промисловості;
РРП	– Реєстр респондентів - підприємств;
FR	– показник запасів палива;
FU	– показник обсягу використаного палива;
CV	– коефіцієнт варіації;
SE	– стандартна похибка;
ME	– гранична похибка;
n	– обсяг вибірки;
N	– загальна кількість одиниць сукупності в основі вибірки;
w	– статистична вага одиниць спостереження.

I. Вступ

1. У Методиці описано порядок формування сукупностей одиниць статистичного спостереження, що вивчаються, які використовують та/або мають у запасах паливо.

Методика розроблена з метою впровадження системного підходу до процесу формування, забезпечення надійності та об'єктивності статистичної інформації щодо використання та запасів палива на основі застосування вибіркового методу відбору одиниць спостереження. Необхідність цього виникла у зв'язку з дією підпункту 1 пункту 1 Закону України "Про захист інтересів суб'єктів подання звітності та інших документів у період дії воєнного стану або стану війни" [1], що унеможлиблює отримання Держстатом статистичної звітності в повному обсязі.

2. Методика ґрунтується на положеннях законів України "Про офіційну статистику" [2], "Про інформацію" [3] та відповідає зобов'язанням України щодо імплементації статті 355 Глави 5 "Статистика" розділу V Угоди про асоціацію з ЄС у частині питань щодо надання вчасних і надійних даних, які можна порівняти на міжнародному рівні.

3. Методика враховує міжнародні стандарти Регламенту (ЄС) № 1099/2008 Європейського Парламенту та Ради від 22 жовтня 2008 року про статистику в галузі енергетики [9].

4. Об'єктом статистичного спостереження за формою 4-мтп (річна) (цільовою сукупністю) є юридичні особи та відокремлені підрозділи юридичних осіб (місцеві одиниці), які використовують та/або мають у запасах паливо.

5. Для проведення процедури формування вибіркової сукупності одиниць статистичного спостереження, що вивчається, використовується обстеження безпосередньо одиниць статистичного спостереження за методом несучільного вивчення сукупностей – методом стратифікації та відсікання.

Вибірковий метод дозволяє отримати репрезентативну інформацію для розрахунків використання та запасів палива на державному рівні.

Державне статистичне спостереження "Використання та запаси палива" за формою № 4-мтп (місячна) "Звіт про використання та запаси палива" (далі – ДСС) здійснюється на постійній основі щомісячно. Щомісячні результати ДСС узагальнюються по Україні в цілому, за регіонами за такими показниками:

запаси палива на кінець звітного місяця – за вісьма видами палива, згідно із НПП (вугілля кам'яне, вугілля буре, нафта сира, газ природний, бензин моторний, газойлі (паливо дизельне), мазути паливні важкі, пропан і бутан скраплені) [10];

обсяг використаного палива за звітний місяць – за вісьма видами палива, згідно із НПП (вугілля кам'яне, вугілля буре, нафта сира, газ природний, бензин

моторний, газойлі (паливо дизельне), мазути паливні важкі, пропан і бутан скраплені) [10].

6. За результатами ДСС рівень надійності оцінювання основних показників має відповідати таким критеріям: на національному рівні коефіцієнт варіації оцінок показників складає не більше 5 %; за стратами за обсягами споживання палива) – не більше 15 %, за окремими показниками на регіональному рівні – не більше 15 %.

Показники щодо обсягів запасів і використаного палива по країні в цілому та по регіонах формуються з урахуванням адміністративних даних ДПС про обсяги реалізованого пального й обсяги залишків пального на кінець звітнього періоду в платників акцизного податку на пальне.

7. Процедури, описані в Методиці, реалізуються в Python 3.11.5.

8. Для цілей цієї Методики терміни вживаються в значеннях, наведених у статтях 1, 8, 9, 10 Закону України "Про офіційну статистику" [2], законах "Про ринок електричної енергії" [4], "Про ринок природного газу" [5], пунктах 2, 5, 11, 14, 15, 20, 26, 39, 42, 52, 56, 60, 67, 72–75 Глосарія до плану статистичного спостереження [6], розділі I Методологічних положень державного статистичного спостереження "Реєстр респондентів" [7] і Керівництві з енергетичної статистики [10].

II. Джерела інформації

Джерелами інформації для проведення ДСС "Використання та запаси палива" є дані, отримані від респондентів:

за формами звітності № 4-мтп (річна) та № 4-мтп (місячна) "Звіт про використання та запаси палива";

адміністративні дані ДПС про обсяги реалізованого пального й обсяги залишків пального на кінець звітнього періоду в платників акцизного податку на пальне (юридичних осіб та фізичних осіб-підприємців) за регіонами України;

інформація ДСС "Реєстр респондентів".

III. Генеральна сукупність (основа вибірки) ДСС

1. Одиницями спостереження ДСС є юридичні особи та відокремлені підрозділи юридичних осіб (місцеві одиниці). Дослідження охоплює вищезазначені одиниці, які використовують та/або мають у запасах паливо.

2. Генеральна сукупність одиниць статистичного спостереження формується на основі даних ДСС "Реєстр респондентів" (у частині СРП) і включає одиниці, які є активними станом на 01 листопада року T та відповідають таким критеріям:

- 1) тип статистичної одиниці: місцева одиниця;
- 2) інституційний сектор економіки підприємства, частиною якого є місцева одиниця, відповідно до KICE [8]: S.11 "Нефінансові корпорації", S.12 "Фінансові корпорації", S.13 "Сектор загального державного управління";
- 3) організаційно-правова форма господарювання за КОПФГ: 110–193, 230–281, 310–390, 410–495, 510–590, 610, 620;
- 4) основний вид економічної діяльності за КВЕД: секції А–R;
- 5) ознака в Реєстрі респондентів щодо наявності ліцензії для здійснення діяльності у сферах енергетики та комунальних послуг;
- 6) ознака в Реєстрі респондентів щодо наявності ліцензії на перевезення пасажирів і вантажів повітряними транспортними засобами;
- 7) ознака в Реєстрі респондентів щодо використання альтернативних джерел енергії у сфері електроенергетики;
- 8) наявність у переліку юридичних осіб - отримувачів пального (за адміністративними даними ДПС).

3. Генеральній сукупності одиниць відповідає сукупність одиниць, що вивчається за формою № 4-мтп (річна), яка є основою вибірки для формування сукупності одиниць, що вивчається за формою № 4-мтп (місячна).

IV. Дизайн вибірки та методичні підходи до її формування

1. Для формування сукупності одиниць, що вивчається за формою № 4-мтп (місячна), застосовується вибірковий метод. Дизайн вибірки відповідає стратифікованій випадковій вибірці одиниць спостереження.

Процедура стратифікації основи вибірки полягає в попередньому розподіленні основи вибірки на регіони з подальшим відбором певної кількості одиниць для безпосереднього обстеження в кожному з регіонів за обсягами використання палива.

2. У межах регіонів формуються такі страти:

сукупність одиниць, які використовують вугілля буре, нафту сиру, включаючи газовий конденсат та мазути паливні важкі (страта обстежуються на суцільній основі);

сукупність одиниць, які мають лише в запасах будь-який із видів палива спостереження (страта обстежуються на суцільній основі);

сукупність нетипових одиниць, які використовують будь-який з п'яти видів палива – вугілля кам'яного, газу природного, бензину моторного, палива дизельного, пропану й бутану скраплених, в обсягах, більших за граничні (страта обстежується на суцільній основі, граничний обсяг використання кожного виду палива для кожного регіону є варіативним та визначається за спеціальною процедурою). Розподіл обсягів використання палива за одиницями генеральної сукупності наведено в додатку 1;

сукупність одиниць, які не мають даних щодо використання та запасів палива (страта обстежуються на суцільній основі);

сукупність нетипових одиниць, які використовують будь-який з п'яти видів палива – вугілля кам'яного, газу природного, бензину моторного, палива дизельного, пропану й бутану скраплених, в обсягах, більших за граничні (страта обстежується на суцільній основі). Граничні обсяги визначаються на основі аналізу дисперсії обсягів використання кожного виду палива для кожного регіону за одиницями: граничному обсягу відповідає мінімальна кількість одиниць для обстеження, що забезпечує встановлений рівень надійності результатів;

сукупність одиниць, які використовують будь-який з п'яти видів палива – вугілля кам'яного, газу природного, бензину моторного, палива дизельного, пропану й бутану скраплених, в обсягах, менших за граничні (страта обстежується на вибірковій основі).

Ілюстрація розподілу кількості одиниць для суцільного й вибіркового обстеження використання пропану та бутану скраплених для Вінницької області наведена в додатку 1.

3. Обсяг вибірки для проведення ДСС визначається з урахуванням умови забезпечення необхідного рівня надійності оцінок показників щодо обсягу використаного палива виду h в регіоні r на кінець звітної періоду FU_{rh} , які вимірюються на вибірковій основі.

За критерій надійності оцінювання показників пропонується використовувати коефіцієнти варіації вибірових оцінок – $CV(FU_{rh})$. Якщо встановлена гранична межа надійності оцінок показників CV_{max} (наприклад, коефіцієнт варіації оцінювання показника становить не більше $CV_{max} = 15\%$: $CV(FU_{rh}) \leq 15\%$), обсяг вибірки визначається за показником щодо використання палива (для кожного виду окремо):

$$n_{rh}(FU_{rh}) = \frac{\sigma^2(FU_{rh})}{\bar{FU}_{0rh}^2 \times (CV_{max})^2 + \sigma^2(FU_{rh})/N_{rh}}, \quad (1)$$

де $n_{rh}(FU_{rh})$ – обсяг вибірки, необхідний для визначення показника FU для палива виду h в регіоні r з установленим рівнем надійності (у цьому випадку з коефіцієнтом варіації 15 %);

$\sigma^2(FU_{rh})$ – дисперсії ознак, за якими визначаються показники, за обстеженими одиницями спостереження;

$$\sigma^2(FU_{rh}) = \frac{\left(\sum_{i=1}^{n_{rh}} (x_{rhi} - \bar{x}_{rh})^2 \right)}{n_{rh} - 1}, \quad (2)$$

де x_{rhi} – значення обсягу використаного палива h для i -ї одиниці вибірки в r -му регіоні;

$\bar{x}_{rh}$ – середнє значення обсягу використаного палива h в r -му регіоні;

$\widehat{FU}_{0rh}$ – очікувані значення показників (визначаються за результатами ДСС попередніх періодів);

N_{rh} – загальна кількість одиниць сукупності в основі вибірки, що використовують паливо виду h в регіоні r .

Формула (1) враховує кінцевий (обмежений) обсяг сукупності одиниць спостереження в основі вибірки по кожній страті. Страта визначається регіоном та видом палива, що використовує підприємство. Приклад розрахунку обсягу вибірки для національного та регіонального рівня наведено в додатку 2.

4. За кожним регіоном визначається загальний обсяг вибірки n_r , що забезпечує оцінювання показників щодо обсягів запасів та використання палива за видами палива з установленим рівнем надійності:

$$n_r = \sum_h n_{rh}. \quad (3)$$

Загальний обсяг вибірки по Україні n визначається шляхом підсумовування обсягів вибірок за всіма регіонами за формулою:

$$n = \sum_r n_r. \quad (4)$$

5. При опрацюванні запропонованої схеми визначення обсягу вибірки необхідно врахувати, що значна кількість страт обстежується на суцільній основі, це підвищує загальний рівень надійності оцінювання показників.

Загальна кількість одиниць для відбору визначається з урахуванням очікуваного рівня участі одиниць в ДСС. Найбільш детальне врахування рівня участі для страт за видами палива r_{rh} є бажаним, але навряд чи доцільним. Кількості одиниць для відбору з урахуванням середньомісячних рівнів участі визначаються за формулами:

$$n'_{rh} = \frac{n_{rh}}{r_r}. \quad (5)$$

6. При організації щомісячних обстежень за умови формування основи вибірки та вибіркової сукупності одиниць один раз на рік необхідно враховувати вибуття одиниць за весь період обстеження. Отже, при розрахунку рівня участі r_r мають ураховуватися кумулятивні кількості вибуття одиниць вибірки за 12

місяців. Такий підхід забезпечує необхідну кількість обстежуваних одиниць на останній дванадцятий місяць обстеження.

7. Загальна кількість підприємств для відбору n' визначається за формулою, аналогічною (4):

$$n' = \sum_r n'_r. \quad (6)$$

При реалізації процедур відбору одиниць спостереження за стратами перелік одиниць, який складає основу вибірки в кожній страті, упорядковується за зменшенням обсягу використаного палива та нумерується. Це забезпечує неявну стратифікацію генеральної сукупності за цією ознакою та підвищує репрезентативність результатів обстеження. Відбір одиниць вибірки здійснюється з упорядкованого переліку одиниць основи вибірки в межах кожної страти на основі процедури систематичного відбору з рівними ймовірностями.

8. За результатами відбору визначаються ймовірності відбору кожної одиниці вибірки. На основі цих ймовірностей розраховуються статистичні ваги обстежених одиниць для поширення результатів обстеження на генеральну сукупність.

V. Методичні підходи до поширення результатів обстеження на генеральну сукупність

1. Для поширення результатів ДСС на генеральну сукупність розраховуються статистичні ваги обстежених одиниць вибірки. При цьому для одиниць вибірки кожної страти визначається базова вага w_{Brh} як величина, обернена до ймовірностей їх відбору:

$$w_{Brh} = \frac{N_{rh}}{n'_{rh}}. \quad (7)$$

Отже, для всіх одиниць вибірки в межах страти базова вага є однаковою. Відповідно до формули (7), якщо з певної страти до обстеження обираються всі одиниці (страти, що обстежуються на суцільній основі), базова вага дорівнює одиниці, $w_{Brh} = 1$.

2. Після проведення обстеження базова вага щомісячно коригується з урахуванням фактичного рівня участі одиниць в обстеженні (фактичної кількості обстежених одиниць з урахуванням причин невідповідей). Коригування здійснюється за формулою:

$$w'_{rh} = w_{Brh} \times \frac{n'_{rh}}{n_{rh} \text{ (факт)}}, \quad (8)$$

де w'_{rh} – кінцева вага одиниць вибірки в r -му регіоні;
 $n_{rh}(\text{факт})$ – фактична кількість обстежених одиниць у страті.

3. Оскільки вибірка одиниць спостереження формується один раз на рік, то коригування ваг за формулою (8) має здійснюватися кожного місяця. При цьому для першого місяця обстеження (для січня) розрахунки здійснюються за формулою, аналогічною (8):

$$w'_{rh(1)} = w_{Brh} \times \frac{n'_{rh}}{n_{rh}(\text{факт},1)}. \quad (9)$$

Для всіх наступних місяців формули будуть такими:

$$w'_{rh(2)} = w'_{rh(1)} \times \frac{n_{rh}(\text{факт},1)}{n_{rh}(\text{факт},2)} = w_{Brh} \times \frac{n'_{rh}}{n_{rh}(\text{факт},2)}. \quad (10)$$

...

$$w'_{rh(12)} = w'_{rh(11)} \times \frac{n_{rh}(\text{факт},11)}{n_{rh}(\text{факт},12)} = w_{Brh} \times \frac{n'_{rh}}{n_{rh}(\text{факт},12)}. \quad (11)$$

VI. Методи компенсації відсутніх даних (імпутації)

1. Пропущені дані, на основі яких розраховуються показники, можуть виникнути внаслідок ненадання або несвоєчасного надання одиницями спостереження інформації, наявності аутлайєрів (викидних значень) тощо.

2. Для заповнення пропущених даних по кожному з видів палива застосовується метод імпутації "найближчого сусіда", за яким замість пропущених даних по певній одиниці сукупності підставляються дані іншої одиниці сукупності – найближчого сусіда. При цьому передбачається, що сусід має бути максимально близьким до одиниці, дані для якої імпутуються, за набором основних характеристик.

Ураховуючи щомісячну періодичність ДСС, як сусід може виступати сама одиниця спостереження з характеристиками за попередній місяць.

3. При практичній реалізації процедур імпутації необхідно дотримуватися загальних рекомендацій, а саме:

1) імпутацію даних доцільно проводити послідовно у відповідності з порядком ознак за збільшенням кількості пропусків у них;

2) процедуру імпутації слід максимально автоматизувати так, щоб втручання фахівця, який здійснює імпутацію, було мінімальним;

3) за можливості слід імпутувати не окремі дані, а групи даних (наприклад, якщо за певною одиницею немає даних щодо використання та запасів певних видів палива, то від "найближчого сусіда" слід узяти одразу всі необхідні значення, а не імпутувати їх по одному незалежно);

4) усі дані, що відновленні за допомогою процедури імпутації, слід позначати так, щоб завжди можна було виявити їх (для цього в масиві даних по одиниці створюється нова ознака (флаг), яка дорівнює "1" для імпутованих значень і "0" для не імпутованих значень).

4. Процедура імпутації автоматизується в середовищі Python (див. додаток 3).

VII. Оцінювання показників за результатами ДСС

1. Показники ДСС за певним видом палива за результатами щомісячного обстеження оцінюються шляхом зважування індивідуальних значень показників обстежених одиниць по кожній страті. Для показника сумарних запасів палива виду h , враховуючи те, що одиниці спостереження обстежуються на суцільній основі, розрахунки здійснюються за формулою:

$$\widehat{FR}_{rh} = \sum_{i=1}^{n_{rh}} y_{rhi}, \quad (12)$$

де $\widehat{FR}_{rh}$ – оцінка сумарного обсягу запасів палива h для r -го регіону;

y_{rhi} – значення показника запасів палива h для i -ї одиниці вибірки в r -му регіоні.

2. Для показника обсягів використаного палива формули є аналогічними:

$$\widehat{FU}_{rh} = \sum_{i=1}^{n_{rh}} w'_{rhi} \times x_{rhi}, \quad (13)$$

де $\widehat{FU}_{rh}$ – оцінка сумарного обсягу використаного палива h для r -го регіону;

x_{rhi} – значення обсягу використаного палива h для i -ї одиниці вибірки в r -му регіоні;

w'_{rhi} – кінцева вага i -ї одиниці вибірки в r -му регіоні.

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

VIII. Визначення характеристик надійності оцінювання показників за результатами обстеження показників вибіркового обстеження

1. Надійність показників, що вимірюються за результатами суцільного спостереження, вважається абсолютною, тобто похибок вибірки для них немає.

2. Для визначення надійності оцінювання показників ДСС, що вимірюються на основі вибіркового методу, в межах кожної страти визначаються стандартні похибки вибірки за формулою:

$$SE(\bar{x}_{rh}) = \frac{\sigma(x_{rh})}{\sqrt{n_{rh}}} \times \sqrt{\left(1 - \frac{n_{rh}}{N_{rh}}\right)}, \quad (14)$$

3. Стандартна похибка оцінювання показника за певною кількістю страт, наприклад регіону, за всіма H видами палива розраховується за формулою:

$$SE(\bar{x}_r) = \sqrt{\sum_{h=1}^H \left(\frac{N_{rh}}{N_h}\right)^2 \times SE^2(\bar{x}_{rh})}. \quad (15)$$

Формула (15) враховує страти, що обстежуються на суцільній основі, для яких стандартна похибка $SE(\bar{x}_{rh})$ дорівнює нулю.

4. Для оцінок сумарних значень показників (агрегованих показників для поширення) стандартна похибка визначається за формулою:

$$SE(\widehat{FU}_{rh}) = N_{rh} \times SE(\bar{x}_{rh}). \quad (16)$$

5. Коефіцієнт варіації оцінювання показників за результатами ДСС визначається за формулою:

$$CV(FU_{rh}) = \frac{SE(\widehat{FU}_{rh})}{\widehat{FU}_{rh}}. \quad (17)$$

6. Для визначення меж довірчих інтервалів для 95 % рівня довірчої ймовірності розраховується гранична похибка вибірки:

$$ME(\widehat{FU}_{rh}) = \pm 1,96 \times SE(\widehat{FU}_{rh}) \quad (18)$$

Додаток 1
до Методики
(пункт 2 розділу IV)

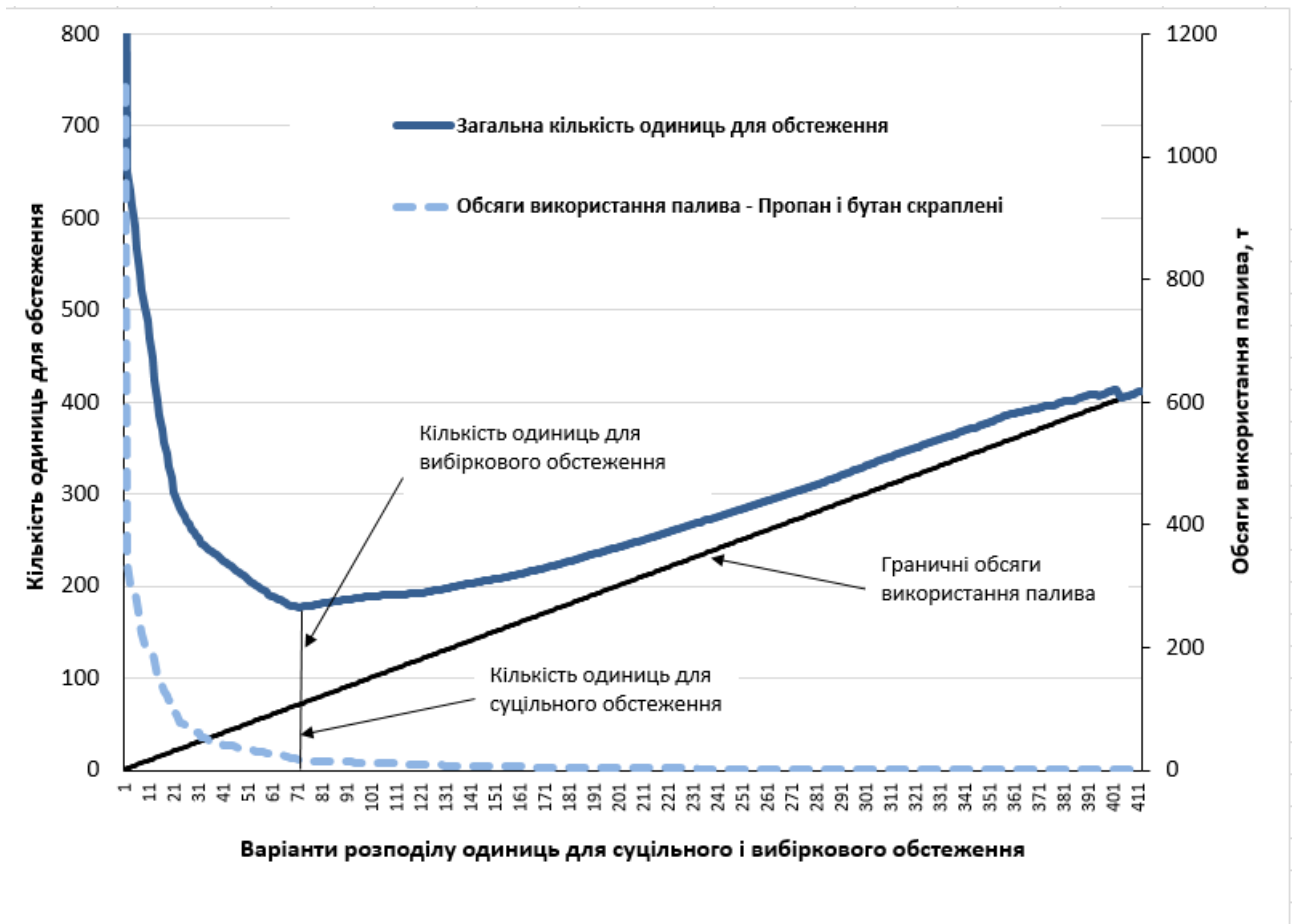


Рисунок 1. Ілюстрація результатів застосування процедури визначення кількості одиниць для суцільного й вибіркового обстеження використання пропану та бутану скраплених по Вінницькій області

Визначення обсягів вибірки для реалізації обстеження за видами використаного палива

Оцінки параметрів, необхідних для визначення обсягу вибірки по п'яти видах палива: вугілля кам'яне, газ природний, бензин моторний, газойлі (паливо дизельне), пропан і бутан скраплені, наведено у табл. 1. Наведені параметри визначено за масивом даних річного обстеження "Використання та запаси палива" 2020 року, наданим для апробації Методики. За встановлений рівень надійності приймається рівень, якому відповідає максимально припустимий коефіцієнт варіації для національного рівня у 5 %.

Таблиця 1. Параметри для визначення обсягу вибірки за видами використаного палива на національному рівні

Вид палива	Вибіркове обстеження				Суцільне обстеження
	Кількість одиниць, N_{rh}	Частка одиниць, $\frac{N_{rh}}{N_h}$	Середнє значення показника, $\bar{F}U_{0rh}$	Дисперсія, $\sigma^2(FU_{rh})$	Кількість одиниць
Вугілля кам'яне	3 121	0,8864	823,17	3621,04	400
Газ природний	15034	0,9288	36,08	8925,03	1152
Бензин моторний	33518	0,9649	6,12	72,94	1219
Газойлі (паливо дизельне)	29047	0,9685	62,08	10830,41	945
Пропан і бутан скраплені	9440	0,9303	5,79	58,05	707
Всього	X				4423

Обсяг вибірки визначається за формулою (1).

Наприклад, для вугілля кам'яного:

$$n_{rh}(FU_{rh}) = \frac{3621,04}{(823,17^2 \times 0,05^2 + 3621,04/3\,121)} = 651 \text{ одиниця}$$

Результати розрахунків для всіх видів використаного палива наведені в табл. 2.

Таблиця 2. Оцінка обсягів вибірки для обстеження одиниць за видами використаного палива на національному рівні

Вид палива	Обсяг вибірки, одиниць
Вугілля кам'яне	651
Газ природний	2319
Бензин моторний	761
Газойлі (паливо дизельне)	1082
Пропан і бутан скраплені	644
Всього	5457

Отже, загальна кількість одиниць для обстеження п'яти видів використання палива становить $5457 + 4423 = 9880$ одиниць, або 20,43 % генеральної сукупності, обсяг якої дорівнює 48360 одиниць.

У таблицях 3, 4 для прикладу наведено результати аналогічних розрахунків для Вінницької області. При цьому за встановлений рівень надійності приймається рівень, якому відповідає максимально припустимий коефіцієнтом варіації у 10 %.

Таблиця 3. Параметри для визначення обсягу вибірки за видами використаного палива для Вінницької області

Вид палива	Вибіркове обстеження				Суцільне обстеження
	Кількість одиниць, N_{rh}	Частка одиниць, $\frac{N_{rh}}{N_h}$	Середнє значення показника, $\hat{F}U_{0rh}$	Дисперсія, $\sigma^2(FU_{rh})$	Кількість одиниць
Вугілля кам'яне	0	0	7061,33	0	214
Газ природний	564	0,8481	16,68	627,85	101
Бензин моторний	1416	0,9568	5,38	44,92	64
Газойлі (паливо дизельне)	1279	0,9726	74,26	11817,56	36
Пропан і бутан скраплені	341	0,8277	4,03	17,09	71
Всього				X	486

Таблиця 4. Оцінка обсягів вибірки для обстеження одиниць за видами використаного палива у Вінницькій області

Вид палива	Обсяг вибірки, одиниць
Вугілля кам'яне	0
Газ природний	161
Бензин моторний	140
Газойлі (паливо дизельне)	184
Пропан і бутан скраплені	81
Всього	566

**Інструктивні матеріали
щодо розрахунку показників щодо використання та запасів палива на
основі адміністративних даних з використанням мови програмування
Python версії 3.11.5**

I. Початок роботи з Python

1. Установіть Python на комп'ютер.

Для встановлення Python на операційну систему Windows (32/64 bit) перейдіть за посиланням <https://www.python.org/downloads/windows/>. На сторінці сайту будуть доступними всі версії Python, оберіть останню стабільну версію Python (на грудень 2024 року – це Python 3.13.1) та завантажте файл інсталяції, який розташований наприкінці сторінки з описом обраної версії Python.

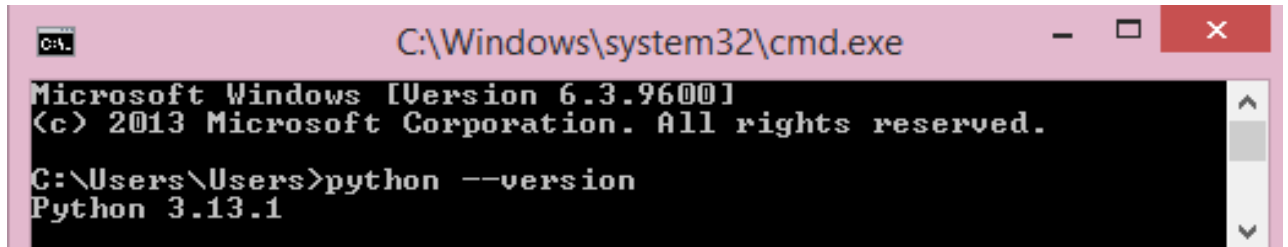
Або скористайтесь посиланням <https://www.python.org/ftp/python/3.13.1/python-3.13.1-amd64.exe> (актуально на грудень 2024 року).

Відкрийте файл інсталяції та дотримуйтесь кроків процесу встановлення. На початку оберіть опцію "Add Python to PATH".



Після встановлення Python на комп'ютер перевірте, чи все встановлено правильно, для цього відкрийте командний рядок (натисніть комбінацію клавіш Win + R, введіть cmd і натисніть Enter) і введіть команду "python –version".

Якщо все встановлено правильно, ви побачите версію Python, яку ви щойно встановили.

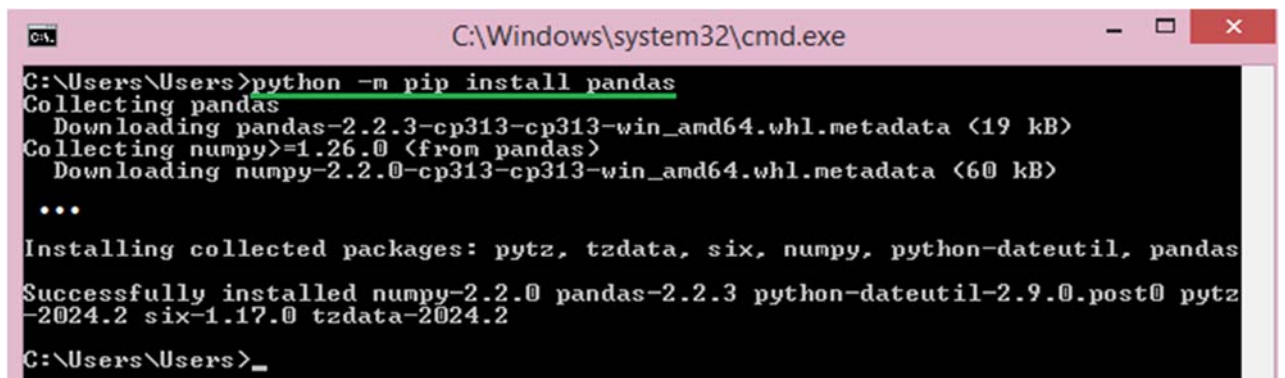


```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.3.9600]
(c) 2013 Microsoft Corporation. All rights reserved.

C:\Users\Users>python --version
Python 3.13.1
```

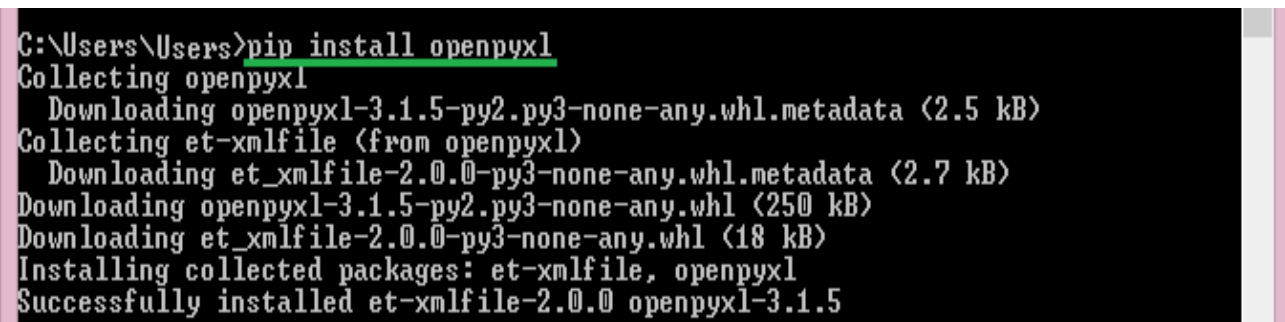
У цьому ж вікні виконайте такі кроки для встановлення пакетів, які будуть використовуватись у подальшому:

1) уведіть команду "python -m pip install pandas". Натисніть "Enter" і дочекайтесь успішного встановлення пакету на ваш комп'ютер



```
C:\Windows\system32\cmd.exe
C:\Users\Users>python -m pip install pandas
Collecting pandas
  Downloading pandas-2.2.3-cp313-cp313-win_amd64.whl.metadata (19 kB)
Collecting numpy>=1.26.0 (from pandas)
  Downloading numpy-2.2.0-cp313-cp313-win_amd64.whl.metadata (60 kB)
...
Installing collected packages: pytz, tzdata, six, numpy, python-dateutil, pandas
Successfully installed numpy-2.2.0 pandas-2.2.3 python-dateutil-2.9.0.post0 pytz-2024.2 six-1.17.0 tzdata-2024.2
C:\Users\Users>_
```

2) уведіть команду "pip install openpyxl". Натисніть "Enter" і дочекайтесь успішного встановлення пакету на ваш комп'ютер.



```
C:\Users\Users>pip install openpyxl
Collecting openpyxl
  Downloading openpyxl-3.1.5-py2.py3-none-any.whl.metadata (2.5 kB)
Collecting et-xmlfile (from openpyxl)
  Downloading et_xmlfile-2.0.0-py3-none-any.whl.metadata (2.7 kB)
Downloading openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)
Downloading et_xmlfile-2.0.0-py3-none-any.whl (18 kB)
Installing collected packages: et-xmlfile, openpyxl
Successfully installed et-xmlfile-2.0.0 openpyxl-3.1.5
```

3) уведіть команду "python -m pip install scipy". Натисніть "Enter" і дочекайтесь успішного встановлення пакету на ваш комп'ютер.

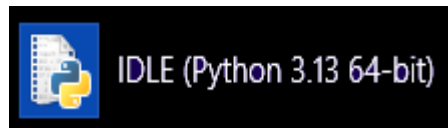
```
C:\Users\Users>python -m pip install scipy
Collecting scipy
  Downloading scipy-1.14.1-cp313-cp313-win_amd64.whl.metadata (60 kB)
Requirement already satisfied: numpy<2.3, >=1.23.5 in c:\users\tanya\appdata\local\programs\python\python313\lib\site-packages (from scipy) (2.2.0)
  Downloading scipy-1.14.1-cp313-cp313-win_amd64.whl (44.5 MB)
----- 44.5/44.5 MB 3.8 MB/s eta 0:00:00
Installing collected packages: scipy
Successfully installed scipy-1.14.1
```

2. Почніть роботу з Python.

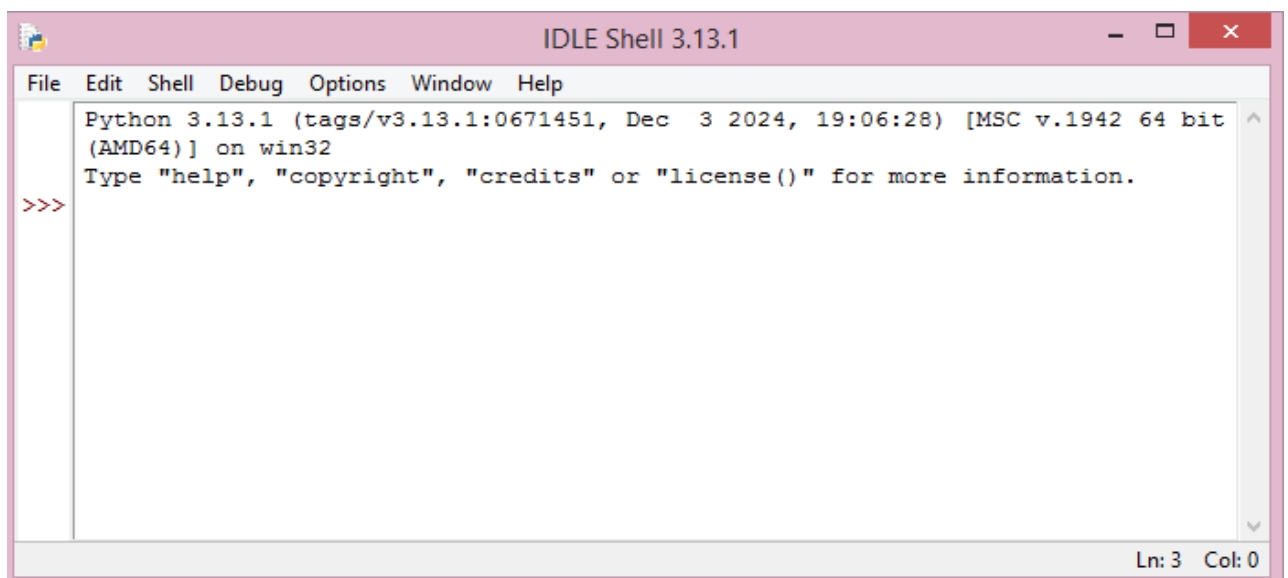
Зазвичай код пишеться в окремих файлах (у будь-якому текстовому редакторі наприклад, Notepad), зберігається з розширенням ".py", а потім запускається через командний рядок Python.

Для більш зручного програмування рекомендуємо використовувати спеціалізовану програму IDLE – це інтегроване середовище розробки, яке встановлюється за замовчуванням разом із Python.

Для початку роботи з кодом програми розрахунків запустіть IDLE.



Відкриється вікно оболонки IDLE.



II. Реалізація розрахунків у Python

Формування переліку одиниць для суцільного та вибіркового обстежень використання підприємствами палива

1. Підготуйте вхідні дані у вигляді таблиці у форматі ".xlsx"., де в першому рядку будуть назви змінних, а в колонках – значення змінних. Збережіть її в робочій теці (у якій у подальшому будуть збережені результати розрахунків). Ця таблиця обов'язково має містити перелічені нижче змінні:

KDMO – індивідуальний код місцевої одиниці;

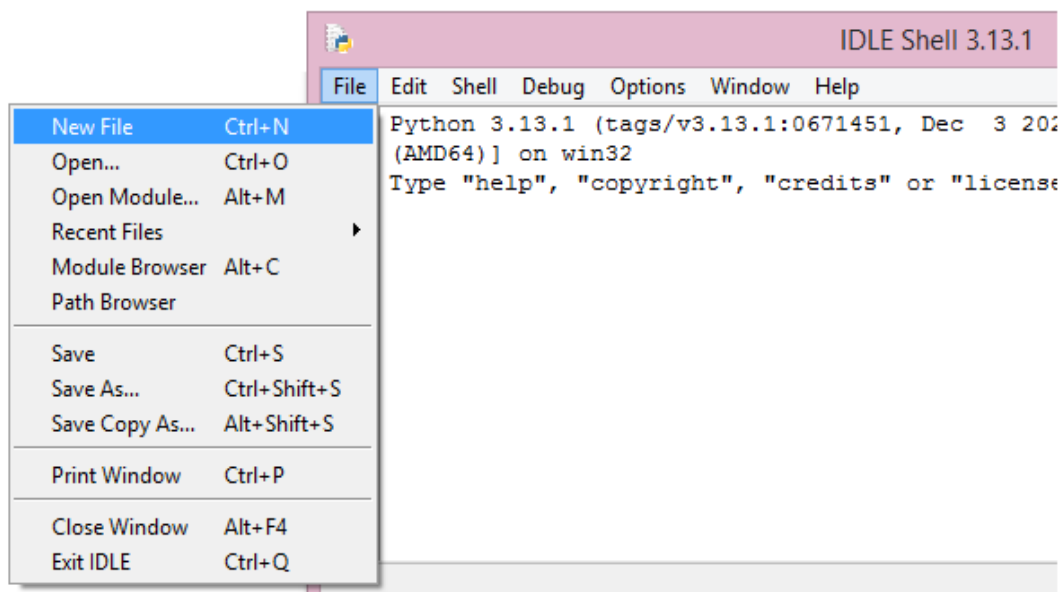
Oblast – код регіону місцезнаходження підприємства;

обсяги використання 9 видів палива: P1110 (Вугілля кам'яне), P1120 (Вугілля буре), P1310 (Газ природний), P1410 (Нафта сира, у т. ч. нафта, одержана з мінералів бітумінозних), P1415 (Газовий конденсат), P1430 (Бензин моторний), P1440 (Газойлі (паливо дизельне)), P1490 (Мазути паливні важкі), P1540 (Пропан і бутан скраплені);

обсяги запасів 9 видів палива: Z1110 (Вугілля кам'яне), Z1120 (Вугілля буре), Z1310 (Газ природний), Z1410 (Нафта сира, у т. ч. нафта, одержана з мінералів бітумінозних), Z1415 (Газовий конденсат), Z1430 (Бензин моторний), Z1440 (Газойлі (паливо дизельне)), Z1490 (Мазути паливні важкі), Z1540 (Пропан і бутан скраплені).

2. Перейдемо до формування переліку одиниць для суцільного та вибіркового обстежень використання підприємствами палива. Щоб реалізувати цю процедуру, виконайте такі кроки: чому далі нумерація латиною?

1) відкрийте вікно редактора.



2) скопіюйте код програми, який наведено в рамці нижче, вставте його в порожнє вікно редактора та збережіть файл у робочу теку.

```
import tkinter as tk
from tkinter import filedialog, messagebox
import pandas as pd
import numpy as np

# Словник для декодування кодів областей
oblast_dict = {
    '01': 'Автономна Республіка Крим',
    '05': 'Вінницька область',
    '07': 'Волинська область',
    '12': 'Дніпропетровська область',
    '14': 'Донецька область',
    '18': 'Житомирська область',
    '21': 'Закарпатська область',
    '23': 'Запорізька область',
    '26': 'Івано-Франківська область',
    '32': 'Київська область',
    '35': 'Кіровоградська область',
    '44': 'Луганська область',
    '46': 'Львівська область',
    '48': 'Миколаївська область',
    '51': 'Одеська область',
    '53': 'Полтавська область',
    '56': 'Рівненська область',
    '59': 'Сумська область',
    '61': 'Тернопільська область',
    '63': 'Харківська область',
    '65': 'Херсонська область',
    '68': 'Хмельницька область',
    '71': 'Черкаська область',
    '73': 'Чернівецька область',
    '74': 'Чернігівська область',
    '80': 'Київ',
    '85': 'Севастополь'
}

# Словник з поясненнями для палива
fuel_descriptions = {
    'P1110': 'Вугілля кам'яне, т',
    'P1310': 'Газ природний, тис.м³',
    'P1430': 'Бензин моторний, т',
    'P1440': 'Газойлі (паливо дизельне), т',
    'P1540': 'Пропан і бутан скраплені, т'
}
```

```

# Функція для додавання пояснень до колонок
def add_fuel_descriptions(df):
    updated_columns = []
    for col in df.columns:
        if col in fuel_descriptions:
            updated_columns.append(f"{col} ({fuel_descriptions[col]})")
        else:
            updated_columns.append(col)
    df.columns = updated_columns
    return df

# Функція для логування
def log(message):
    log_text.insert(tk.END, f"{message}\n")
    log_text.see(tk.END)
    root.update()

# Функція для обчислення обсягу вибірки та кроку
def calculate_sample(data, CV=0.05):
    # Видаляємо NaN
    filtered_data = data[~np.isnan(data)]
    if len(filtered_data) == 0:
        log(" - Усі дані містять NaN або є порожніми. Пропускаємо.")
        return 1 # Мінімальний крок

    # Розрахунок середнього значення та дисперсії
    mean_value = np.mean(filtered_data)
    variance = np.var(filtered_data)
    log(f" - Середнє значення: {mean_value:.4f}")
    log(f" - Дисперсія: {variance:.4f}")

    # Розрахунок обсягу вибірки
    n = variance / ((mean_value * CV) ** 2)
    n = min(n, len(filtered_data)) # Обмеження: обсяг вибірки не може перевищувати
    кількість даних
    log(f" - Обсяг вибірки: {n:.2f}")

    # Розрахунок кроку вибірки
    h = len(filtered_data) / n
    log(f" - Розрахований крок вибірки: {h:.4f}")
    return h

# Функція для обробки файлу
def process_file():
    try:
        # Завантаження файлу
        file_path = filedialog.askopenfilename(filetypes=[("Excel files", "*.xlsx")])

```

```

if not file_path:
    return
log("Завантаження даних...")
df = pd.read_excel(file_path, dtype={'KDMO': str})

# Заміна текстових 'NaN' на справжні NaN
df.replace('NaN', np.nan, inplace=True)
log(" - Усі текстові 'NaN' замінено на справжні NaN.")

# Фільтрування областей
oblast_codes = df['Oblast'].unique()
output_file = "Processed_Data.xlsx"
writer = pd.ExcelWriter(output_file, engine='openpyxl')
for oblast_code in oblast_codes:
    oblast_name = oblast_dict.get(str(oblast_code).zfill(2),
f"Область_{oblast_code}")
    log(f"Обробка області: {oblast_name}...")
    oblast_df = df[df['Oblast'] == oblast_code]

    # Генеральна сукупність
    oblast_df.to_excel(writer, sheet_name=f"{oblast_name} - Генеральна",
index=False)
    log(f" - Генеральна сукупність збережена для області: {oblast_name}.")
    result = []
    for fuel_col in ['P1110', 'P1310', 'P1430', 'P1440', 'P1540']:
        log(f"Обробка палива: {fuel_col}...")

        # Фільтрація даних: виключаємо NaN, некоректні та нульові значення
        fuel_data = oblast_df[['KDMO', fuel_col]].dropna(subset=[fuel_col]).copy()
        fuel_data[fuel_col] = pd.to_numeric(fuel_data[fuel_col], errors='coerce')
        fuel_data = fuel_data[fuel_data[fuel_col] > 0]
        # Логування
        log(f" - Записів до очищення: {len(oblast_df)}")
        log(f" - Записів після очищення: {len(fuel_data)}")

        # Перевірка, чи дані є після очищення
        if fuel_data.empty:
            log(f" - Немає даних для палива: {fuel_col}. Пропускаємо обробку.")
            continue

        # Спеціальна логіка для 80-P1430
        if oblast_code == 80 and fuel_col == "P1430":
            log(" - Використовується спеціальна логіка для 80 області - паливо
P1430.")

        # Фільтрація даних: виключаємо NaN, некоректні та нульові значення
        fuel_data = oblast_df[['KDMO', fuel_col]].dropna(subset=[fuel_col]).copy()

```

```

fuel_data[fuel_col] = pd.to_numeric(fuel_data[fuel_col], errors='coerce')
fuel_data = fuel_data[fuel_data[fuel_col] > 0]

# Логуювання
log(f" - Записів до очищення: {len(oblast_df)}")
log(f" - Записів після очищення: {len(fuel_data)}")

# Перевірка, чи дані є після очищення
if fuel_data.empty:
    log(f" - Немає даних для палива: {fuel_col}. Пропускаємо обробку.")
    continue

# Суцільна вибірка (5 % найбільших значень)
full_sample = fuel_data.head(int(len(fuel_data) * 0.05)).copy()
full_sample['no_sampling'] = 1
full_sample['sampling'] = 0
full_sample_count = len(full_sample)

# Вибіркова сукупність (решта 95 %)
remaining_data = fuel_data.iloc[int(len(fuel_data) * 0.05):].copy()
if not remaining_data.empty:
    # Обчислення параметрів для 95 % даних
    mean_value = remaining_data[fuel_col].mean()
    variance = remaining_data[fuel_col].var()
    sample_size = max(1, int(variance / ((mean_value * 0.05) ** 2)))
    # Обсяг вибірки
    h = max(1, len(remaining_data) / sample_size) # Розрахунок кроку
    first_step = int(h * np.random.uniform(0, 1)) # Перший випадковий крок

    # Вибір записів за кроком
    sampled_data = remaining_data.iloc[first_step::max(1, int(h))].copy()
    sampled_data['no_sampling'] = 0
    sampled_data['sampling'] = 1
    sampled_count = len(sampled_data)

    # Додавання до результатів
    result.append(full_sample)
    result.append(sampled_data)

    # Логуювання
    log(f" - Середнє значення (95 %): {mean_value:.4f}")
    log(f" - Дисперсія (95 %): {variance:.4f}")
    log(f" - Обсяг вибірки: {sample_size}")
    log(f" - Крок вибірки: {h:.2f}")
    log(f" - Кількість записів у суцільній сукупності: {full_sample_count}")
    log(f" - Кількість записів у вибірковій сукупності: {sampled_count}")
else:

```

```

log(" - Дані для вибіркової вибірки відсутні.")
result.append(full_sample)

continue # Пропускаємо стандартну логіку для цієї пари
# Сортуння даних у спадному порядку
fuel_data = fuel_data.sort_values(by=fuel_col, ascending=False).copy()

# Умова для малої кількості записів
if len(fuel_data) < 20:
    fuel_data['no_sampling'] = 1
    fuel_data['sampling'] = 0
    result.append(fuel_data)
    log(f" - Всі записи додано до суцільної вибірки (менше 20 записів).")
    continue

# Суцільна вибірка (30 % найбільших значень)
full_sample = fuel_data.head(int(len(fuel_data) * 0.3)).copy()
full_sample['no_sampling'] = 1
full_sample['sampling'] = 0
full_sample_count = len(full_sample)

# Вибіркова вибірка (решта 70 %)
remaining_data = fuel_data.iloc[int(len(fuel_data) * 0.3):].copy()
if not remaining_data.empty:
    h = calculate_sample(remaining_data[fuel_col].values)
    first_step = int(h * np.random.uniform(0, 1)) # Перший випадковий крок
    sampled_data = remaining_data.iloc[first_step::max(1, int(h))].copy()
    # Обираємо записи за кроком
    sampled_data['no_sampling'] = 0
    sampled_data['sampling'] = 1
    sampled_count = len(sampled_data)

# Додавання до результатів
result.append(full_sample) # Додаємо суцільну вибірку
result.append(sampled_data) # Додаємо тільки обрані записи з вибіркової
вибірки

# Логування
log(f" - Кількість записів у суцільній сукупності: {full_sample_count}")
log(f" - Кількість записів у вибірковій сукупності: {sampled_count}")
else:
    log(f" - Дані для вибіркової вибірки відсутні.")
    result.append(full_sample)

# Об'єднання результатів із сортуванням
if result:
    final_sample = pd.concat(result).sort_index() # Об'єднуємо результати

```

```

# Переміщуємо колонки no_sampling та sampling після P1540
cols = final_sample.columns.tolist()
cols.remove('no_sampling')
cols.remove('sampling')
p1540_index = cols.index('P1540') # Отримуємо індекс останньої колонки
палива
cols.insert(p1540_index + 1, 'no_sampling') # Додаємо після P1540
cols.insert(p1540_index + 2, 'sampling')
final_sample = final_sample[cols]

# Додаємо пояснення до колонок
final_sample = add_fuel_descriptions(final_sample)

# Записуємо у файл
final_sample.to_excel(writer, sheet_name=f"{oblast_name} - Вибірки",
index=False)
log(f"Результати для області {oblast_name} збережено.")

# Додавання до результатів
result.append(full_sample)
result.append(sampled_data)

writer.close()
log("Результати збережено у файл: Processed_Data.xlsx")
messagebox.showinfo("Готово", f"Результати збережено у файл: {output_file}")

except Exception as e:
    messagebox.showerror("Помилка", f"Сталася помилка: {e}")
    log(f"Помилка: {e}")

# Головне вікно
root = tk.Tk()
root.title("Обробка файлу")

# Текстове поле для логів
log_text = tk.Text(root, wrap=tk.WORD, height=20, width=80)
log_text.pack(pady=10)

# Кнопка завантаження файлу
load_button = tk.Button(root, text="Завантажити файл", command=process_file)
load_button.pack(pady=10)

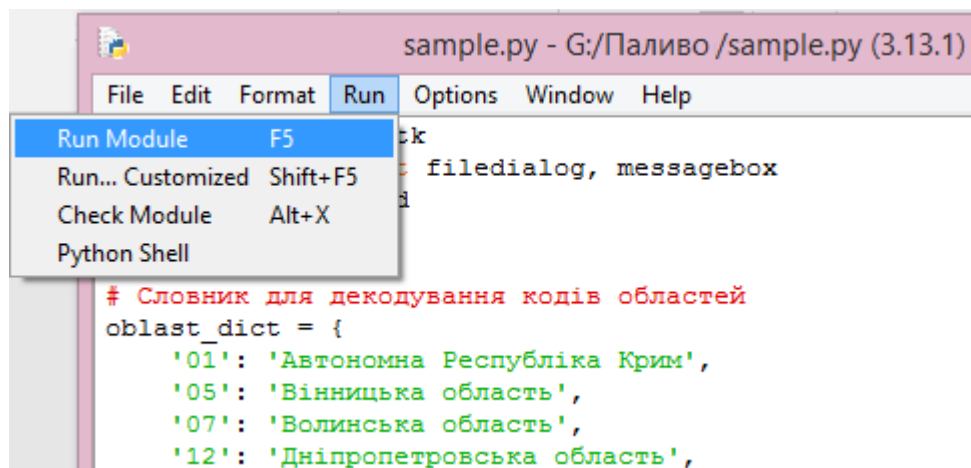
# Кнопка копіювання логів
def copy_logs():
    root.clipboard_clear()
    root.clipboard_append(log_text.get("1.0", tk.END).strip())
    root.update()
    messagebox.showinfo("Скопійовано", "Текст логів скопійовано у буфер обміну.")

```

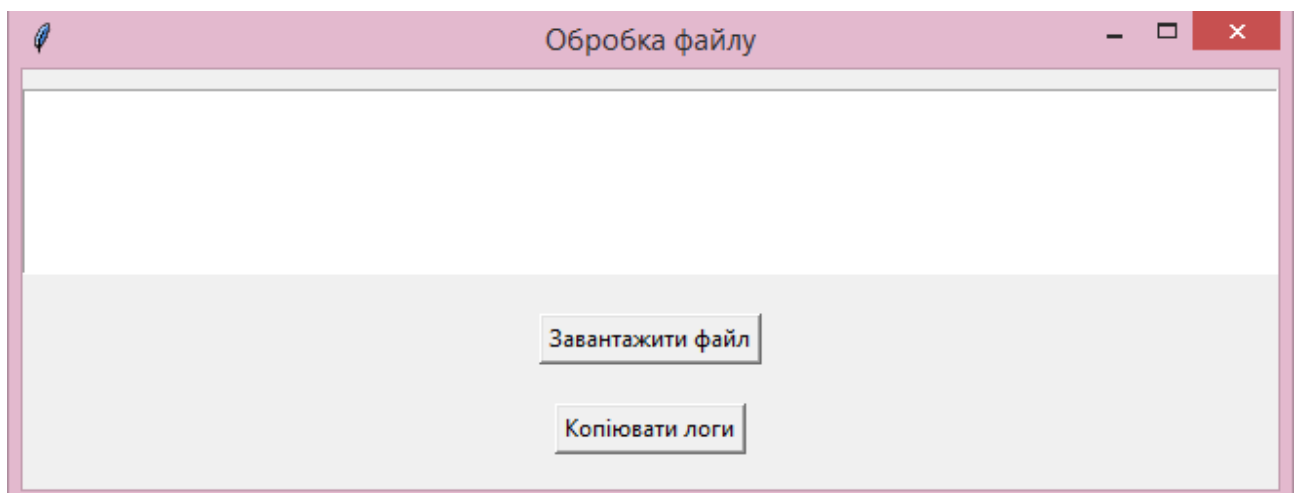
```
copy_button = tk.Button(root, text="Копіювати логи", command=copy_logs)
copy_button.pack(pady=10)

# Запуск вікна
root.mainloop()
```

3) запустіть код програми, обравши на панелі меню "Run" → "Run Module" або натисніть на клавішу F5.

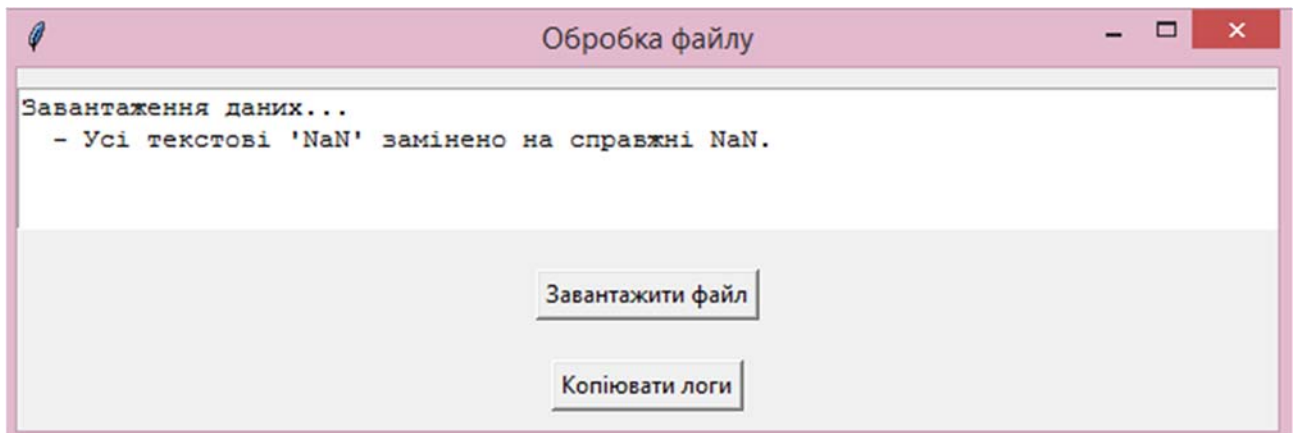


У результаті відкриється вікно програми формування переліку одиниць сукупностей для суцільного та вибіркового обстежень використання підприємствами палива.

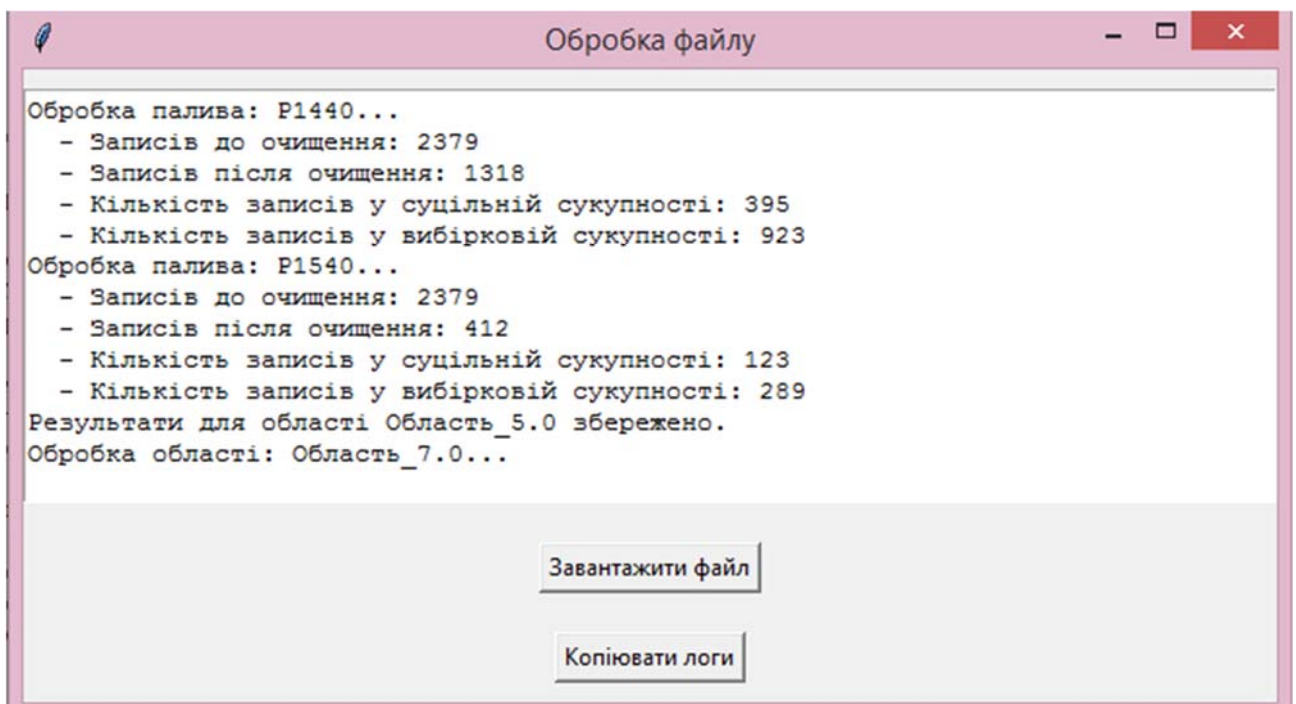


4) натисніть на клавішу "Завантажити файл" і оберіть файл із вхідними даними, який було створено за описом в пункті 1.

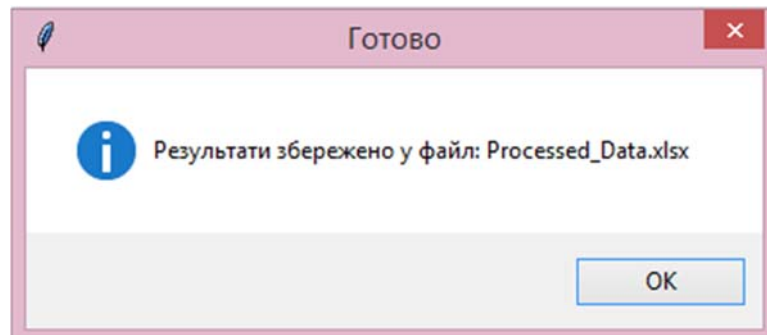
Почнеться обробка файлу.



У процесі обробки ви зможете спостерігати, як у кожному регіоні та по кожному виду палива обчислюється кількість підприємств, щодо яких наявна інформація про використання та запаси палива, а також їхня кількість у сукупностях для суцільного та вибіркового обстеження.



Коли ці розрахунки буде виконано, програма повідомить про збереження результатів у файл "Processed_Data.xlsx", який буде збережений у вашій робочій теці.



**Приклад таблиці вхідних даних щодо обсягів використання
одиницями спостереження визначених видів палива**

KDMO	Регіон	P1110 (Вугілля кам'яне, т)	P1310 (Газ природний, тис.м³)	P1430 (Бензин моторний, т)	P1440 (Газойлі (паливо дизельне), т)	P1540 (Пропан і бутан скраплені, т)
369420620001	80				3797,9	
369420620001	80			3556		
369420620001	80					204
1353900378	80			3251		
1353900378	80				5332,5	
401089810001	80			1737		
401089810001	80				250,7	
401089810001	80		165,7			
401085830001	80			917,3		
401085830001	80				358,3	
427954900001	80			874,2		
427954900001	80		819018,8			
427954900001	80				1586,8	
427954900001	80					215,6
313167180001	80				40878,6	
313167180001	80					42,7
313167180001	80			867,5		
302653380001	80			830,7		
302653380001	80				455,3	
377252850001	80				133,7	
377252850001	80			738,3		
377252850001	80		2197,9			
...	...	...	...	...	...	..

**Приклад вихідної таблиці роботи програми, які формуються
по кожній області окремо**

KDMO	Регіон	P1110	P1310	P1430	P1440	P1540	Суцільне обстежен- ня	Вибіркове обстежен- ня	Агрегований код типів палива, що викорис- товуються підприєм- ством
126350001	80			6,6	10,1			1	110
135400001	80			18,4	2,2			1	110
136980001	80			10,9	4,4			1	110
153320001	80			1,2				1	100
156220001	80			17	13,2			1	110
321060082	80		2583,5	0,2	4			1	1110
321120001	80			154,7	12		1		110
321290643	80			146,6	337,6		1		110
340510001	80			237,4	20,5		1		110
341860001	80			61,6	1,1	0,2	1		111
414670001	80			5,6				1	100
460770001	80			4,5				1	100
639280001	80			0,5				1	100
653240001	80			17	10,5			1	110
1002270001	80						1		0
1002270003	80						1		0
1002270004	80			80,3	25,6	10,8	1		111
1093390001	80			4,5	1,5	0,4		1	111
1107340001	80						1		0
1139680001	80			10,5	7,5			1	110
1314000001	80						1		0
1324570001	80			18,3	41,8	9,3		1	111
1353900018	80			56,4	582,7		1		110
1353900378	80			3251	5332,5	3,2	1		111

5) наступним кроком після формування вибіркової сукупності є формування систематичної випадкової вибірки. У середовищі Python застосовується функція, аргументами якої є масив даних вибіркової сукупності "data" та значення бажаного обсягу вибірки "n". Ця функція виконує такі дії:

- одиниці генеральної сукупності упорядковуються за визначеним порядком;
- автоматично обирається випадкова початкова точка;
- автоматично визначається крок відбору;
- формується перелік одиниць відбору.

```

import random
def systematic_random_sample(data, n):
    """
    Selects a systematic random sample of size n from the data.
    Parameters:
    data (list): The list of items to sample from.
    n (int): The desired sample size.
    Returns:
    list: A systematic random sample of size n.
    """

    # Determine the step size
    step = len(data) // n

    # Randomly choose a starting point between 0 and step-1
    start = random.randint(0, step - 1)

    # Select every step-th element starting from the random start
    sample = [data[i] for i in range(start, len(data), step)]

    # Return only the first n items, as there may be extras if the data length
    # is not perfectly divisible by n
    return sample[:n]

# Example usage
data = list(range(1, 101)) # A dataset of numbers 1 to 100
n = 10 # Desired sample size

sample = systematic_random_sample(data, n)
print("Systematic Random Sample:", sample)

```

Крок відбору обчислюється шляхом ділення загальної кількості одиниць даних на обсяг вибірки. Випадковий початок (перший крок) визначається множенням випадкового числа (від 0 до 1) на крок відбору. Одиниці для обстеження визначаються додаванням кроку відбору до першого кроку.

Застосувавши цю функцію до вибіркової сукупності для кожного регіону та за кожним видом палива, ви отримаєте вибірку для відповідної групи підприємств.

За результатами відбору визначаються ймовірності відбору кожної одиниці вибірки та її статистична вага.

III. Програма визначення статистичної надійності оцінок показників

1. За результатами обстеження підготуйте вхідні дані у вигляді таблиці, яка міститиме значення трьох змінних KDMO: індивідуальний код місцевої одиниці,

зміна обсягу використання/ запасів палива (по якій будуть розраховані характеристики статистичної надійності) та вагу підприємства.

Збережіть таблицю в робочій теці у форматі ".csv" і назвіть цей файл "data.csv".

2. Щоб реалізувати процедуру визначення статистичної надійності оцінок показників, виконайте такі кроки:

1) відкрийте вікно редактора;

2) скопіюйте код програми, який наведений у рамці нижче, вставте його в порожнє вікно редактора та збережіть файл у робочу теку;

3) запустіть код програми, обравши на панелі меню "Run" → "Run Module" або натисніть на клавішу F5.

```
import numpy as np # Бібліотека для роботи з масивами та числовими обчисленнями
import scipy.stats as stats # Модуль для статистичних розрахунків
import csv # Модуль для роботи з CSV-файлами

def calculate_sample_statistics(sample_data, confidence_level=0.95):
    """
    Розрахунок основних статистичних показників вибірки.

    Параметри:
        sample_data (list of lists): Масив даних, кожен рядок має формат [номер,
        значення, коефіцієнт].
        confidence_level (float): Рівень довіри (за замовчуванням 95 %).

    Повертає:
        dict: Словник з розрахованими статистичними величинами:
            - Стандартна помилка
            - Гранична похибка
            - Довірчий інтервал
            - Коефіцієнт варіації
    """
    # Перетворення масиву даних у NumPy-масив для обчислень
    sample_values = np.array([row[1] for row in sample_data]) # Беремо тільки числові
    значення другого стовпця
    sample_weights = np.array([row[2] for row in sample_data]) # Використовуємо
    коефіцієнти (ваги)
    n = len(sample_values) # Розмір вибірки (кількість елементів)
    sample_mean = np.mean(sample_values) # Середнє значення вибірки
    sample_std = np.std(sample_values, ddof=1) # Стандартне відхилення (з поправкою
    для вибірки)

    # Розрахунок стандартної помилки (SE)
```

```

standard_error = sample_std / np.sqrt(n)
# Z-значення для заданого рівня довіри
z_score = stats.norm.ppf((1 + confidence_level) / 2) # Наприклад, для 95 % Z ≈ 1.96

# Розрахунок граничної похибки
margin_of_error = z_score * standard_error
# Визначення довірчого інтервалу
confidence_interval = (sample_mean - margin_of_error, sample_mean +
margin_of_error)

# Розрахунок коефіцієнта варіації
coefficient_of_variation = (sample_std / sample_mean) * 100 if sample_mean != 0
else np.nan

# Оцінка середнього значення з урахуванням ваг
weighted_mean = np.sum(sample_values * sample_weights) /
np.sum(sample_weights)

# Оцінка сумарного значення з урахуванням ваг
weighted_sum = np.sum(sample_values * sample_weights)
# Повернення результатів у вигляді словника
return {
    "Standard Error": standard_error,
    "Margin of Error": margin_of_error,
    "Confidence Interval": confidence_interval,
    "Coefficient of Variation": coefficient_of_variation,
    "Weighted Mean": weighted_mean,
    "Weighted Sum": weighted_sum
}

# Основний блок програми: читання файлу та обчислення
filename = "data.csv" # Назва файлу, який містить дані
sample_data = [] # Порожній список для зберігання даних із файлу

try:
    # Відкриття файлу в режимі читання
    with open(filename, "r") as file:
        reader = csv.reader(file) # Ініціалізація CSV-рідера для обробки файлу
        next(reader) # Пропускаємо заголовок файлу, якщо він є

    # Читання даних із файлу рядок за рядком
    for row in reader:
        # Перетворення рядків у числові значення: [номер, значення, коефіцієнт]
        sample_data.append([
            int(row[0]), # Перший стовпець - номер (ціле число)
            float(row[1].replace(",", ".")), # Другий стовпець - значення (дробове число)
            float(row[2].replace(",", ".")) # Третій стовпець - коефіцієнт (дробове число)
        ])

```

```

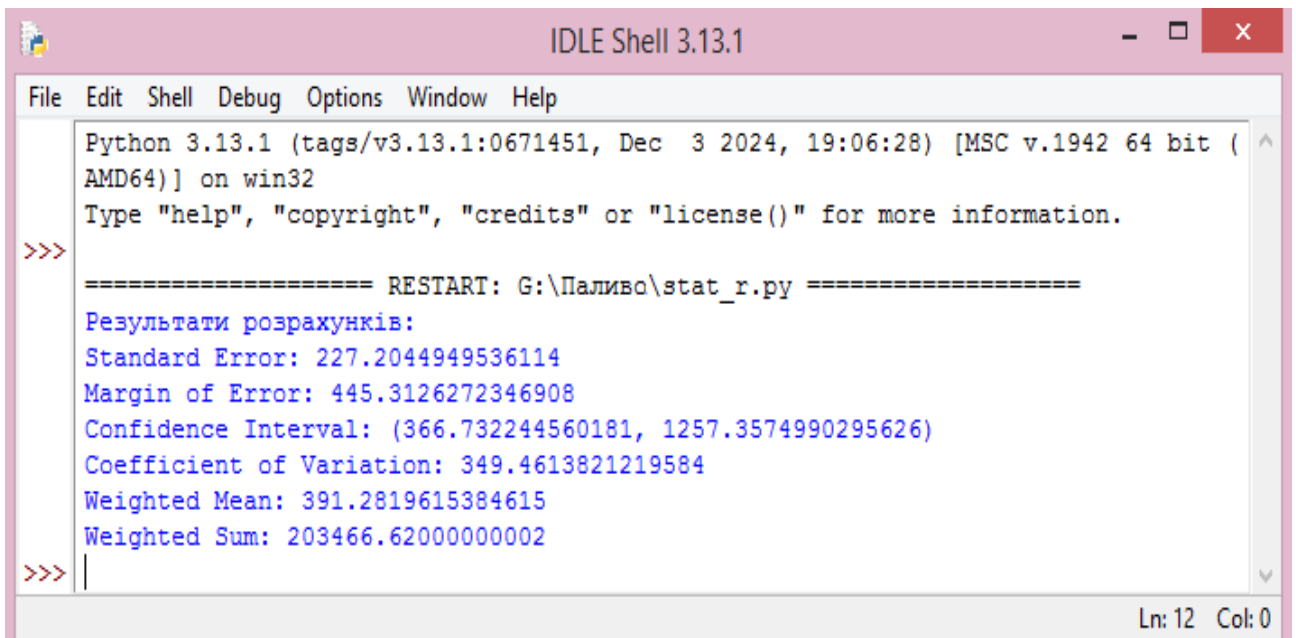
# Виклик функції для обчислення статистичних показників
results = calculate_sample_statistics(sample_data)

# Виведення результатів
print("Результати розрахунків:")
for stat, value in results.items():
    if isinstance(value, tuple): # Якщо значення - це кортеж, як у довірчому інтервалі
        value = tuple(float(v) for v in value) # Перетворюємо всі елементи на float
    elif isinstance(value, np.float64): # Якщо це число типу np.float64
        value = float(value) # Перетворюємо на звичайний float
    print(f"{stat}: {value}")

except FileNotFoundError:
    # Обробка випадку, якщо файл не знайдено
    print(f"Файл '{filename}' не знайдено!")
except Exception as e:
    # Загальна обробка помилок
    print(f"Виникла помилка: {e}")

```

У результаті буде отримано характеристики надійності, як-от: стандартна похибка, гранична похибка, довірчий інтервал, коефіцієнт варіації. А також оцінка зваженого середнього та зваженого сумарного обсягу використання/запасів певного виду палива.



```

IDLE Shell 3.13.1
File Edit Shell Debug Options Window Help
Python 3.13.1 (tags/v3.13.1:0671451, Dec 3 2024, 19:06:28) [MSC v.1942 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: G:\Паливо\stat_r.py =====
Результати розрахунків:
Standard Error: 227.2044949536114
Margin of Error: 445.3126272346908
Confidence Interval: (366.732244560181, 1257.3574990295626)
Coefficient of Variation: 349.4613821219584
Weighted Mean: 391.2819615384615
Weighted Sum: 203466.62000000002
>>>
Ln: 12 Col: 0

```

Приклад результатів роботи програми

Область	Оцінка показника використання палива	P1110	P1310	P1430	P1440	P1540
		Вугілля кам'яне, т	Газ природний, тис.м³	Бензин моторний, т	Газойлі (паливо дизельне), т	Пропан і бутан скраплені, т
м. Київ	Середнє значення (Weighted Mean)	70.63	821.13	14.21	71.89	3.48
	Сумарне значення (Weighted Sum)	318903.30	3707416.10	64161.10	324566.00	15710.50
Дніпропетровська область	Середнє значення (Weighted Mean)	2353.37	918.42	9.40	122.95	8.98
	Сумарне значення (Weighted Sum)	8603932.00	3357732.90	34371.20	449492.10	32846.40

IV. Модуль імпутації пропущених значень у середовищі Python

Для імпутації пропущених значень у середовищі Python застосовується функція `KNNImputer` (визначає відсутні значення, ураховуючи значення найближчих сусідів для відсутніх точок даних), яка має такий вигляд:

```
import numpy as np
import pandas as pd
from sklearn.impute import KNNImputer

# Example dataset with missing values
data = {
    'Feature1': [1, 2, np.nan, 4, 5],
    'Feature2': [2, np.nan, 3, 8, 7],
    'Feature3': [np.nan, 4, 6, 9, np.nan]
}
# Convert the data into a pandas DataFrame
df = pd.DataFrame(data)

# Create a KNNImputer object
imputer = KNNImputer(n_neighbors=2) # Use 2 nearest neighbors for imputation
```

```
# Perform the imputation
imputed_data = imputer.fit_transform(df)

# Convert the imputed array back to a pandas DataFrame
imputed_df = pd.DataFrame(imputed_data, columns=df.columns)

# Display the imputed data
print("Original DataFrame with Missing Values:")
print(df)

print("\nImputed DataFrame:")
print(imputed_df)
```

KNNImputer уключає:

процедуру `scikit`, яка заповнює відсутні значення на основі даних для найближчих сусідів;

параметр `n_neighbors`, який контролює, скільки сусідів розглядається для імпутації відсутніх значень;

функцію `fit_transform`, яка заповнює відсутні значення в даних на основі даних для найближчих сусідів.

Список використаних джерел

1. Закон України "Про захист інтересів суб'єктів подання звітності та інших документів у період дії воєнного стану або стану війни" від 03 березня 2022 року № 2115-IX.

URL: <https://zakon.rada.gov.ua/laws/show/2115-20#Text>

2. Закон України "Про офіційну статистику" від 16 серпня 2022 року № 2524-IX.

URL: <https://zakon.rada.gov.ua/laws/show/2524-20#Text>

3. Закон України "Про інформацію" від 02 жовтня 1992 року № 2657-XII.

URL: <https://zakon.rada.gov.ua/laws/show/2657-12#Text>

4. Закон України "Про ринок електричної енергії" від 13 квітня 2017 року № 2019-VIII.

URL: <https://zakon.rada.gov.ua/laws/show/2019-19#Text>

5. Закон України "Про ринок природного газу" від 09 квітня 2015 року № 329-VIII.

URL: <https://zakon.rada.gov.ua/laws/show/329-19#Text>

6. Глосарій до плану статистичного спостереження, затверджений наказом Держкомстату від 29 грудня 2009 року № 498.

URL: https://ukrstat.gov.ua/metod_polog/glos.htm

7. Методологічні положення державного статистичного спостереження "Реєстр респондентів статистичних спостережень", затверджені наказом Держстату від 30 листопада 2021 року № 298 (у редакції наказу Держстату від 16 серпня 2024 року № 207).

URL: https://www.ukrstat.gov.ua/norm_doc/2021/298/298.pdf

8. Кодифікатор адміністративно-територіальних одиниць та територій територіальних громад, затверджений наказом Міністерства розвитку громад та територій від 26 листопада 2020 року № 290 (в редакції наказу Міністерства розвитку громад та територій України від 12 січня 2021 року № 3).

URL: https://www.ukrstat.gov.ua/klasf/nac_kls/tab_kato.htm

9. Regulation (EC) № 1099/2008 of the European Parliament and of the Council of 22 October 2008 on energy statistics

URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=%20CELEX:32008R1099&from=EN>

10. Energy Statistics Manual, OECD/IEA, 2004.

URL: <https://iea.blob.core.windows.net/assets/67fb0049-ec99-470d-8412-1ed9201e576f/EnergyStatisticsManual.pdf>