



Огляд процедур оцінки та обробки даних національної лісової інвентаризації

Артем Щербина

Київ, листопад 2022



About the Project “Sustainable Forestry Implementation” (SFI)

The project “Technical Support to Forest Policy Development and National Forest Inventory Implementation” (SFI) is a project established in the framework of the Bilateral Cooperation Program (BCP) of the Federal Ministry of Food and Agriculture of Germany (BMEL) with the Ministry of Environment and Natural Resources of Ukraine (MENR). It is a continuation of activities started in the forest sector within the German-Ukrainian Agriculture Policy Dialogue (APD) forestry component.

The Project is implemented based on an agreement between GFA Group, the general authorized executor of BMEL, and SFRA since October 2021. On behalf of GFA Group, the executing agencies - Unique land use GmbH and IAK Agrar Consulting GmbH - are in charge of the implementation jointly with SFRA.

The project aims to support sustainable forest management planning in Ukraine and has a working focus on the results in the Forest Policy and National Forest Inventory.

Author

A. Shcherbina

Disclaimer

This paper is published with assistance of SFI but under the solely responsibility of the author A.Shcherbina under the umbrella of the Sustainable Forestry Implementation (SFI). The whole content, particularly views, presented results, conclusions, suggestions or recommendations mentioned therein belong to the authors and do not necessarily coincide with SFI's positions.

Contacts

Troitska Str. 22-24,
Irpin, Kyiv region
+38 (067) 964-77-02

ЗМІСТ

1. Список термінів та скорочень	2
2. Національна інвентаризація лісів	2
2.1 Визначення лісу	2
2.2 Вибіркова сукупність та одиниці вибірки	3
2.3 Одиниці виміру	3
2.4 Вибіркова ділянка	3
3. Зберігання даних	4
4. Моделювання даних	5
4.1 Моделювання висоти	5
4.2 Моделювання об'єму	6
5. Статистична оцінка	7
5.1 Підсумкова оцінка та її точність	8
5.2 Оцінка співвідношень	8
5.3 Пропущені значення	10
5.4 Використання показників	11
6. Потік виконання	11
6.1 Завантаження таблиць	12
6.2 Обробка даних	12
6.3 Обчислення таблиць	13
6.4 Приклади оцінок площ	13
6.5 Багаторічні інвентаризації	14
7. Перелік питань для подальшого аналізу	15
Література	15

1 Список термінів та скорочень

Cov	коваріація двох випадкових величин
Var	дисперсія випадкової величини
E	математичне сподівання випадкової величини
s	площа окремої ділянки
CSV	значення, розділені комами
ДВН	діаметр на висоті грудей (1,3 м)
ІІД	незалежні та однаково розподілені
Деревостан	частина лісу, що відрізняється за сукупністю характеристик (різновид лісових страт)
Запис	одиничний запис у вибірці
Змінна	числове значення, пов'язане з елементом популяції або стратами
Значення	популяційний параметр - функція змінних, що вивчаються
НІЛ	Національна інвентаризація лісів
Популяція	популяція лісу - сукупність скінченної кількості дерев на даній лісовій площі
Підгрупа	субпопуляція, частина популяції, ділянка відповідно до значень факторів
Стратифікація ..	поділ площі (лісу) на окремі підплощі (страти)
Тракт	група з 4-х ділянок
Фактор	категорична характеристика дерева або деревостану, яка використовується для поділу загальних результатів
Характеристика .	категоріальне значення, пов'язане з елементом або стратами сукупності

2 Національна інвентаризація лісів

Національна інвентаризація лісів (НІЛ) - це система щорічних вибірових обстежень земель лісового фонду країни.

2.1 Визначення лісу

Визначення лісу, що використовується для НІЛ:

1. мінімальна площа лісової ділянки: 0,1 га;
2. мінімальне покриття крони (або еквівалентне запасу): 30 відсотків;
3. мінімальна висота дерев у віці стиглості: 5 м;
4. мінімальна ширина лісу: 20 м.

2.2 Вибіркова сукупність та одиниці вибірки

Випадковий відбір дерев з генеральної сукупності замінюється на (випадковий) вибірку ділянок з заданої площі. На карту території країни наноситься сітка квадратів розміром 5,0 на 5,0 км. Кожен інвентаризаційний квадрат містить кластер з чотирьох інвентаризаційних пробних площ, які розглядаються як інвентаризаційний тракт. Тракт розміщується випадковим чином в межах інвентаризаційного квадрату. Кожне інвентаризаційний тракт та пробна площа віднесено до однієї з п'яти часових панелей, як показано на малюнку 1. Запропонована рамка вибірки відповідає інтенсивності щорічної вибірки у розмірі 20 відсотків від загальної кількості трактів, враховуючи тривалість циклу інвентаризації 5 років. Крім того, щорічні інвентаризації можуть розглядатися як окремі вибірки, розглядатися як окремі вибірки.



Рис. 1: Рамка вибірки НІЛ

2.3 Одиниці виміру

Окреслена основа вибірки визначає стратифіковану кластерну випадкову вибірку з комірками квадратної сітки як окремими стратами, інвентаризаційними ділянками як кластерами та інвентаризаційними ділянками як одиницями вибірки. Незважаючи на те, що відстань між пробними площами прийнята таким чином, що вони не повинні потрапляти в один і той самий ліс, деякі характеристики часто корелюють на рівні тракту (порода, тип лісу, вік, порушення тощо) і розгляд їх як незалежних спостережень може призвести до упереджених оцінок. Таким чином, основною розрахунковою одиницею для оцінки параметрів НІЛ обрано обрано тракт.

2.4 Вибіркова ділянка

В НІЛ використовується вибірка з фіксованою круговою ділянкою. Вибіркова ділянка складається з набору кругових ділянок (вкладених ділянок) для вимірювання дерев різного діаметру як показано на малюнку 2 та в таблиці 1.

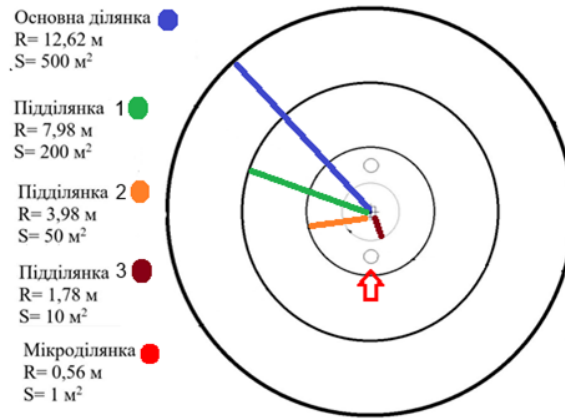


Рис. 2: Приклад дизайну ділянки

При використанні ділянок з фіксованою площею радіуси визначають функцію вибірки дерева. Функція вибірки дерева не є стохастичним процесом, а правилом асоціації між точкою вибірки (центром ділянки) та деревами. Однак, оскільки ми припускаємо, що розташування центру ділянки є випадковим результатом стохастичного процесу, який генерує точки на площі популяції, то вибір дерев є стохастичним процесом [2].

Для розрахунку характеристики дерев зважуються на площу ділянки за допомогою множника, який визначається як $M_s = A_0/A_s$, $s = 0, \dots, 3$, де A_s - площа вкладеної ділянки. Наразі мікроділянки не використовуються для обробки.

Ділянка	Радіус, m	Площа, m^2	Умова	Множник, M_s
(Головна) ділянка 0	12.62	500	$DBH \geq 14$	1
Ділянка 1	3.98	50	$DBH \in [6, 14)$	10
Ділянка 2	1.78	10	$DBH \in [2, 6)$	50
(Мікро) ділянка 3	0.56	1	$DBH \leq 2$ або $H \leq 1.3$	0

Табл. 1: Умови вимірювання дерева на пробній ділянці

Всі лісові ділянки використовуються для збору даних польовими бригадами.

При польовому обстеженні площа кожної пробної площі також розбивається на сегменти (підділянки), що відносяться до різних категорій лісових (нелісових) земель. Таким чином, в межах ділянки наносяться межі різних лісових насаджень. Дереву відносяться до різних підділянок, відповідно до свого розташування.

Перелік змінних та характеристик, що підлягають збору на пробних площах визначено Інструкцією з проведення НІЛ.

3 Зберігання даних

База даних НІЛ створюється у програмі SmallForest в операційній системі Android (рисунок 3). База даних включає групи характеристик для різних об'єктів:

- ділянка: GPS координати, адміністративна інформація, рельєф;
- сегменти (підділянки): інформація про користувача, опис деревостану (яруси, догляд, порушення, індекс ділянки, тип лісу, індекс ділянки), підлісок та чагарники, трав'яна рослинність, сухостій, ґрунт;

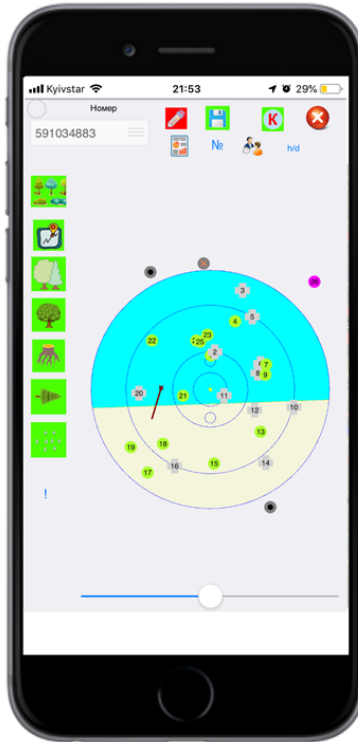


Рис. 3: Програма SmallForest для НІЛ

та змінні для:

1. підрахунку дерев: місцезнаходження, порода, обміри (ДВН); стан (загибель, пошкодження), оцінки (якість стовбура, класи крафту) і т.д.;
2. модельні дерева: обміри (висота, вік), оцінка (стан крони), стан (якість крони), стан (якість стовбура);
3. Лісовідновлювальні дерева: вимірювання (висота), стан (пошкодження);
4. пні: діаметр, стан (гнилість) тощо.

Річні бази даних зберігаються окремо, але об'єднуються для загальних оцінок за кілька років або циклічно для розрахунку приростів.

Структура таблиць бази даних представлена в Додатку 1 до цього документу.

4 Моделювання даних

З багатьох змінних, що спостерігаються для кожного дерева ділянки, тільки порода та діаметр (ДВН) доступні для всіх дерев у вибірці. Висота дерева (Н) вимірюється для вибірки дерев другого етапу (модельних), а об'єм дерев не заміряється. Ці величини повинні бути оцінені на основі наявних вимірювань. Для цього ми застосовуємо дві моделі: для висоти та для об'єму. Вони базуються на таблицях об'ємних сортиментів для дерев [3]. Ці таблиці надають дані про рівень росту, ДВН, висоту та об'єм для 14 видів порід.

4.1 Моделювання висоти

Висота оцінюється за ДВН за допомогою наступної функції

$$H = (a + b * G) * \arctan(c * DBH), \quad (1)$$

де a, b та c - параметри, що підбираються для кожної породи окремо. Наприклад, для однієї з порід значення даних та підібрані криві представлені на малюнку 4. Розрахункові значення параметрів наведені в таблиці 2.

параметр	a	b	c	d	e
значення	10.3	2.20	0.0617	0.42	0.999

Табл. 2: Підібрані значення параметрів для породи 2

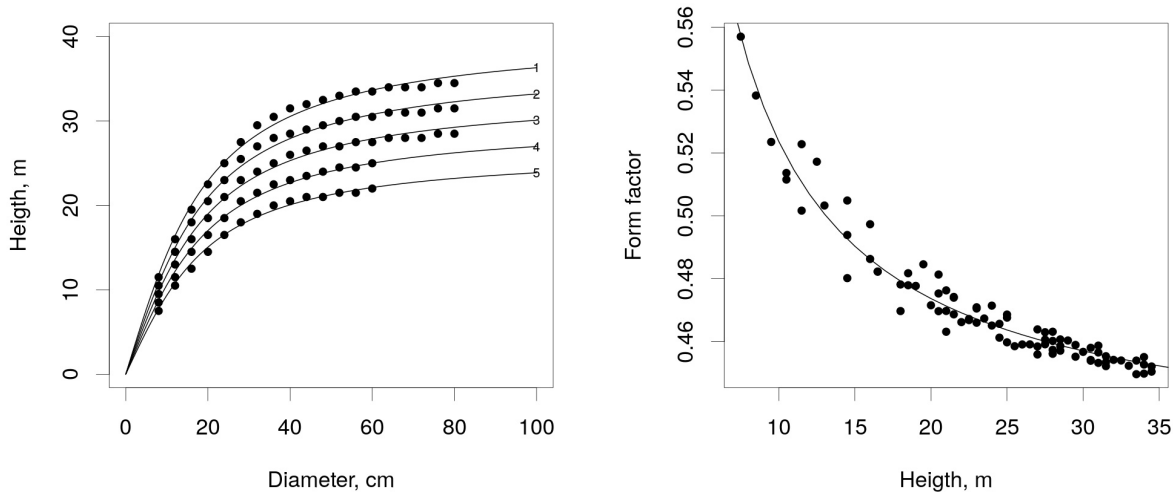


Рис. 4: Моделі висоти та форм-фактору для породи 2

Для застосування моделі ми повинні оцінити рівень росту кожного деревостану. Ми визначаємо окремий деревостан за допомогою комбінації IDPlots, IDSub_Plots та Species. Для всіх дерев моделі ми можемо обчислити рівень приросту за формулою

$$G = \left(\frac{H}{\arctan(c * DBH)} - a \right) / b.$$

Якщо деревостан має декілька модельних дерев, ми можемо усереднити їх значення, щоб отримати рівень приросту цього деревостану. Якщо деревостан не має модельних дерев, ми використовуємо обчислений рівень приросту з опису деревостану опису деревостану.

У випадку, якщо ми все ще не можемо отримати рівень приросту, ми прирівнюємо його до мінімального мінімальному значенню для поточної підділянки.

Використовуючи рівень росту деревостану, ми можемо використати рівняння 1 для обчислення висоти всіх дерев.

Цей процес проілюстровано на малюнку 5 ліворуч з 17 деревами та 7 модельними деревами.

У вибірці Сумської області 2008 року залежність між модельною висотою та прогнозованою висотою проілюстровано на малюнку 5 праворуч.

4.2 Моделювання об'єму

Об'єм дерева описується рівнянням

$$V = f * \pi \left(\frac{DBH}{200} \right)^2 * H, \quad (2)$$

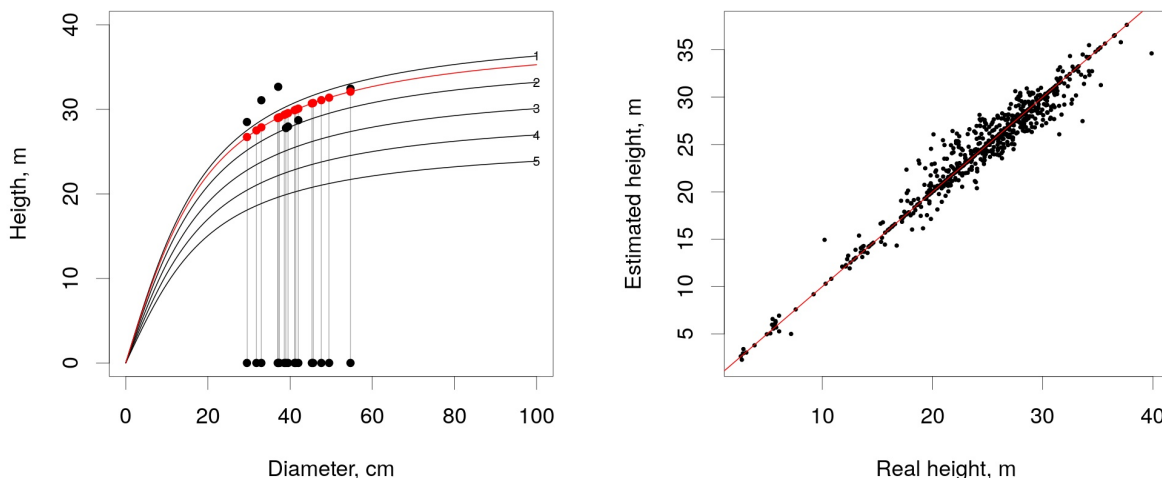


Рис. 5: Оцінка висоти

де f - форм-фактор, тобто характеризує форму дерева і є співвідношенням між об'ємом циліндра та об'ємом дерева.

Використовуючи таблиці об'ємних сортиментів, ми можемо розрахувати форм-фактор за формулою

$$f = \frac{V}{\pi(DBH/200)^2 * H}. \quad (3)$$

Для прогнозування форм-фактора використовується наступна формула

$$f = d + \frac{e}{H},$$

де d та e - параметри, підібрані для кожної породи окремо. Для породи 2 дані та підібрана криві представлені на малюнку 4. Розрахункові значення параметра наведені в таблиці 2.

Середні квадратичні похибки для моделей висоти та об'єму наведені в таблиці 3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
height mse	0.70	4.09	0.76	1.23	2.58	0.89	0.20	3.02	0.53	3.64	0.29	1.44	0.16	0.44
volume mse×100	2.57	1.07	2.56	2.13	1.20	1.68	1.32	2.63	1.59	1.48	1.42	1.31	2.50	3.02

Табл. 3: Середньоквадратичні похибки для моделей для різних порід

5 Статистична оцінка

Провівши інвентаризацію, нам потрібен спосіб оцінити сумарне значення певної характеристики X та її точність для всієї сукупності із загальною площею S . Нехай наша вибірка складається з N трактів площею $s = 4\pi r^2$ кожен. Кожна ділянка має N_i , $i = 1, \dots, N$ записів. Значеннями нашої характеристики є X_{ij} , $j = 1, \dots, N_i$. Тоді, позначимо сумарне значення для тракту i наступним чином

$$X_i = \sum_{j=0}^{N_i} X_{ij}, \quad i = 1, \dots, N.$$

Далі будемо вважати, що величини X_i є незалежними та однаково розподіленими (IID).

5.1 Підсумкова оцінка та її точність

Звичайні незміщені оцінки середнього значення та дисперсії мають вигляд

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i,$$
$$S_{\bar{X}}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2.$$

Оцінка для сумарного значення має вигляд

$$\hat{X} = \frac{S}{s} \bar{X}. \quad (4)$$

Оскільки $S_{\hat{X}}^2 = S_{\bar{X}}^2/N$, то оцінка дисперсії сумарного значення має вигляд

$$S_{\hat{X}}^2 = \frac{S^2}{s^2} S_{\bar{X}}^2 = \frac{S^2}{s^2 N} S_X^2.$$

Для оцінки точності використаємо коефіцієнт варіації (CV). Він визначається наступним чином

$$\mathbf{CV} X = \frac{\sqrt{\mathbf{Var} X}}{\mathbf{E} X}$$

і оцінюється за формулою

$$\widehat{\mathbf{CV}} \hat{X} = \widehat{\mathbf{CV}} \bar{X} = \frac{S_X}{\bar{X} \sqrt{N}}.$$

Оскільки число N зазвичай досить велике (>1000), то можна наближено вважати розподіл \hat{X} рівним $N(\hat{X}, S_{\hat{X}}^2)$. Тоді, ми можемо вибрати довірчу ймовірність $0.5 < \alpha < 1$. Позначимо $C_\alpha = Q^{N(0,1)}((1 + \alpha)/2)$ квантиль стандартного нормального розподілу рівня $(1 + \alpha)/2$. Використовуючи його, можна побудувати довірчий інтервал

$$[\hat{X}(1 - C_\alpha \widehat{\mathbf{CV}} \hat{X}), \hat{X}(1 + C_\alpha \widehat{\mathbf{CV}} \hat{X})].$$

Зазвичай використовують $\alpha = 0.95$, тоді $C_\alpha \approx 1.96$.

Зауваження. При побудові статистичних оцінок одна з найголовніших характеристик є саме незміщеність. Вона гарантує, що зі зростанням об'єму вибірки, оцінка буде необмежено наближатися до справжнього значення. Якщо ж використовується зміщена оцінка, то це твердження не є дійсним. Тобто, зі зростанням об'єму вибірки, оцінка буде наближатися до іншого значення, яке не співпадає зі справжнім. В такому випадку неможна використовувати довірчі інтервали отриманих оцінок.

Одним з прикладів зміщених оцінок є використання інших множників M_s у таблиці 1.

5.2 Оцінка співвідношень

Нехай маємо дві випадкові величини X та Y і ставимо за мету оцінити їх відношення $R = X/Y$. Тут і далі символи \mathbf{E} , \mathbf{Var} та \mathbf{Cov} позначають математичне сподівання, дисперсію та коваріацію випадкових величин. У нашому випадку випадковими величинами є деякі характеристики трактів. Позначимо нормовані змінні $X_n = X/\mathbf{E} X$, $Y_n = Y/\mathbf{E} Y$ та $R_n = X_n/Y_n$. При цьому було показано в [1], наближеннями першого порядку для середнього та дисперсії є:

$$\mathbf{E} R_n \approx \frac{\mathbf{E} X_n}{\mathbf{E} Y_n} = 1,$$
$$\mathbf{Var} R_n \approx \mathbf{Var} X_n - 2 \mathbf{Cov}(X_n, Y_n) + \mathbf{Var} Y_n.$$

Таким чином, оцінки для ненормованої змінної мають вигляд

$$\begin{aligned} \mathbf{E} R &\approx \mathbf{E} R_n \frac{\mathbf{E} X}{\mathbf{E} Y} = \frac{\mathbf{E} X}{\mathbf{E} Y}, \\ \mathbf{CV} R &= \frac{\sqrt{\mathbf{Var} R}}{\mathbf{E} R} = \sqrt{\mathbf{Var} \left(\frac{R}{\mathbf{E} R} \right)} = \sqrt{\mathbf{Var} \left(\frac{R_n}{\mathbf{E} R_n} \right)} \\ &\approx \sqrt{\mathbf{Var} R_n} \approx \sqrt{\mathbf{Var} X_n - 2 \mathbf{Cov}(X_n, Y_n) + \mathbf{Var} Y_n} \\ &= \sqrt{\frac{\mathbf{Var} X}{(\mathbf{E} X)^2} - 2 \frac{\mathbf{Cov}(X, Y)}{\mathbf{E} X \mathbf{E} Y} + \frac{\mathbf{Var} Y}{(\mathbf{E} Y)^2}}. \end{aligned}$$

Нехай для X та Y ми спостерігаємо їх трактові значення $X_i, Y_i, i = 1, \dots, N$. Тоді ми можемо оцінити значення у попередніх формулах відповідно до наступної таблиці:

Value	Estimate	Formula
$\mathbf{E} X$	\bar{X}	$\frac{1}{N} \sum_{i=1}^N X_i$
$\mathbf{E} Y$	\bar{Y}	$\frac{1}{N} \sum_{i=1}^N Y_i$
$\mathbf{Var} X$	S_X^2	$\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$
$\mathbf{Var} Y$	S_Y^2	$\frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$
$\mathbf{Cov}(X, Y)$	S_{XY}^2	$\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})$

Вибіркові оцінки середнього та коефіцієнта варіації для R мають вигляд:

$$\bar{R} \approx \frac{\bar{X}}{\bar{Y}}, \quad (5)$$

$$\widehat{\mathbf{CV}} R \approx \sqrt{\frac{S_X^2}{\bar{X}^2} - 2 \frac{S_{XY}^2}{\bar{X}\bar{Y}} + \frac{S_Y^2}{\bar{Y}^2}} = \sqrt{\frac{1}{N} \left(\frac{S_X^2}{\bar{X}^2} - 2 \frac{S_{XY}^2}{\bar{X}\bar{Y}} + \frac{S_Y^2}{\bar{Y}^2} \right)}. \quad (6)$$

В останній формулі присутній коефіцієнт $1/N$, оскільки ми оцінюємо коефіцієнт варіації відношення підсумків, а не для окремих спостережень з масиву.

Розглянемо приклад значення в таблиці 4.

тракт	підділянка	висота	площа	висота×площа
1	1	10	500	5000
1	2	15	200	3000
1	3	20	300	6000
2	1	20	500	10000
2	2	15	500	7500
3	1	25	300	7500
4	1	20	500	10000
4	2	20	500	10000
4	3	15	300	4500
4	4	25	500	12500

Табл. 4: Виміряні значення для підділянок

У нашому випадку X_i - це загальна висота×площа тракту i , а Y_i - це загальна площа тракту i . Використовуючи таблицю 5 можна обчислити середні значення $\bar{X} = 76000/4 = 19000$, $\bar{Y} = 4100/4 =$

Тракт	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	14000	1000	-5000	-25	25000000	625	125000
2	17500	1000	-1500	-25	2250000	625	37500
3	7500	300	-11500	-725	132250000	525625	8337500
4	37000	1800	18000	775	324000000	600625	13950000
Всього	76000	4100	0	0	483500000	1127500	22450000

Табл. 5: Похідні значення для трактів

1025, дисперсії та коваріацію $S_X^2 = 483500000/3 \approx 161166667$, $S_Y^2 = 1127500/3 \approx 375833$, $S_{XY}^2 = 22450000/3 \approx 7483333$.

Підставивши ці значення у формули (5 - 6) отримаємо отримаємо

$$\bar{R} \approx \frac{37000}{1800} \approx 18.54,$$

$$\widehat{CV} R \approx \sqrt{\frac{1}{4} \left(\frac{161166667}{19000^2} - 2 \frac{7483333}{19000 \times 1025} + \frac{375833}{1025^2} \right)} \approx 0.094$$

5.3 Пропущені значення

Під час збору даних значення деяких змінних можуть бути пропущені або їм можуть бути присвоєні некоректні значення з різних причин. Пізніше це може призвести до некоректних або зміщених оцінок. В ідеалі, помилкові значення повинні бути виявлені та виправлені під час польового контролю. Встановлені процедури контролю повинні перевіряти узгодженість та повноту даних. Для вирішення проблеми пропущених значень можна використовувати декілька підходів.

Одним з найпростіших і найпопулярніших методів є імплікація середнього значення. В цьому методі записи з відсутніми значеннями X_{ij} замінюються середніми значеннями по вибірці. Позначимо пропущене значення через NA.

Якщо припустити, що значення пропущені абсолютно випадково (MCAR), то незміщена оцінка для середнього буде

$$\bar{X}^M = \frac{1}{N - N_M} \sum_{i=1}^N \sum_{X_{ij} \neq \text{NA}} X_{ij},$$

де N_M - загальна кількість пропущених значень

$$N_M = \sum_{i=1}^N \sum_{j=1}^{N_i} \mathbf{1}_{\{X_{ij}=\text{NA}\}}.$$

Імпутація цих значень замість NA дає скориговані значення X_{ij}^M . Тоді ми можемо обчислити загальні значення у трактах як

$$X_i^M = \sum_{j=1}^{N_i} X_{ij}^M.$$

Ці значення можуть бути використані для обчислення вибіркової дисперсії

$$S_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i^M - \bar{X})^2.$$

5.4 Використання показників

Часто нас цікавить характеристика X у частині генеральної сукупності (яка називається доменом), що визначається умовним показником I . Її вибіркові значення мають вигляд I_{ij} , $i = 1, \dots, N$, $j = 1, \dots, N_i$. Якщо ми визначимо нову характеристику $X_{ij}^I = X_{ij}I_{ij}$, то сумарні значення трактів для області матимуть вигляд

$$X_i^I = \sum_{j=0}^{N_i} X_{ij}^I, \quad i = 1, \dots, N.$$

Аналогічно, результати (4) можна застосувати для оцінювання середніх і загальних значень та CV:

$$\bar{X}^I = \frac{1}{N} \sum_{i=1}^N X_i^I, \quad (7)$$

$$S_{\bar{X}^I}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i^I - \bar{X}^I)^2, \quad (8)$$

$$\hat{X}^I = \frac{S}{s} \bar{X}^I, \quad (9)$$

$$\widehat{CV} \hat{X}^I = \frac{S_{\bar{X}^I}}{\bar{X}^I}. \quad (10)$$

Наприклад, показник лісистості може бути використаний для оцінки загальної лісистості лісистості.

Зазвичай побудовані таблиці містять оцінки не тільки для всієї сукупності, але й різні оцінки для значень того чи іншого фактора. В загальному випадку для різних областей можна використовувати декілька факторів. Нехай маємо K -факторну таблицю ($K = 1, 2$ або 3). Нехай значення факторів мають вигляд $F_{ij}^k \in \{C_1^k, C_2^k, \dots, C_{N_k}^k\}$, $k = 0, \dots, K$. Для кожної комбінації значень факторів (f_1, \dots, f_K) визначається індикаторна змінна I

$$I_{ij} = \mathbf{1}_{\{F_{ij}^1=f_1\}} \cdot \dots \cdot \mathbf{1}_{\{F_{ij}^K=f_K\}}.$$

У нашій інвентаризації кожна таблиця поділена на один, два або три фактори. Нижче наведено значення індикаторів для таблиці з двома факторами.

	$F_{ij}^2 = C_1^2$	$F_{ij}^2 = C_2^2$...	$F_{ij}^2 = C_{N_2}^2$
$F_{ij}^1 = C_1^1$	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}} \mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}} \mathbf{1}_{\{F_{ij}^2=C_2^2\}}$...	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}} \mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$
$F_{ij}^1 = C_2^1$	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}} \mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}} \mathbf{1}_{\{F_{ij}^2=C_2^2\}}$...	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}} \mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$
...
$F_{ij}^1 = C_{N_1}^1$	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}} \mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}} \mathbf{1}_{\{F_{ij}^2=C_2^2\}}$...	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}} \mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$

Наприклад, такими факторами можуть бути $F^1 = \text{сосна}$ та $F^2 = \text{Державна власність}$.

6 Потік виконання

Змінні та характеристики, зібрані в лісі, проходять декілька перетворень для отримання вихідних таблиць. Ці етапи описані нижче.

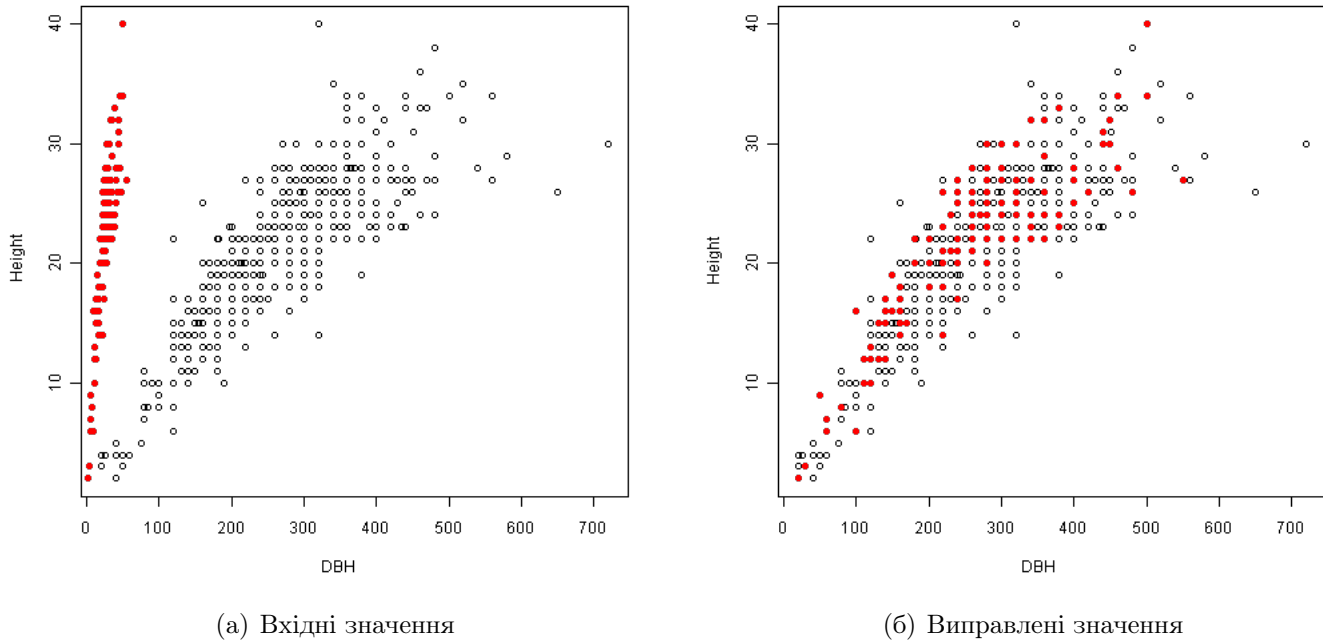


Рис. 6: Коригування вимірювань DBH

6.1 Завантаження таблиць

Єдиний набір даних інвентаризації зберігається в окремій папці, кожна таблиця представлена файлом у форматі csv (значення, розділені комами). Перелік використовуваних на даний момент таблиць зі змінними та характеристиками, що представляють інтерес наведено у файлі `inputTables.csv`.

Оскільки для однієї і тієї ж території послідовно проводиться кілька річних інвентаризацій, ми можемо об'єднати їх дані для підвищення точності оцінок. Відповідно до прийнятої просторово-часової схеми оцінки за п'ятирічний період періоду дають результати для одного кола НЛ.

6.2 Обробка даних

Нижче коротко описані дії, що були здійснені при обробці даних.

1.checkData.r. Для отримання інформації про якість даних виконується декілька перевірок. Наприклад, характеристичні значення повинні знаходитися в межах визначених діапазонів, категоріальні значення повинні бути із заздалегідь визначеного набору, важливі характеристики не повинні бути пропущені.

2.correctData.r. Тут виправлені деякі типові помилки, додані деякі фіктивні значення. Поширеною помилкою є вимір DBH в сантиметрах замість міліметрів. На рисунку 6 зафарбовані точки позначають помилкові вимірювання, які визначаються відношенням між діаметром дерева та його висотою. Після множення DBH на 10 вони стають правильними.

3.moveData.r. Виміряні значення, які згадуються в одній таблиці, переносяться переміщуються до іншої таблиці.

4.prepareIndicators.r. Обчислюються показники для відбору доменів.

5.prepareData.r. Обчислюються деякі додаткові значення, включаючи рівень зростання та оцінки зросту.

6.prepareVolume.r. Об'єм дерева моделюється за допомогою DBH, висоти та рівня росту. Після цього об'єми дерев агрегуються до підділянок. Під час цього процесу об'єми компенсуються коефіцієнтом M_s малих ділянок. Таким чином змінні обсягів для підділянок називаються `Volume_Count`.

7.prepareGroups.r. Обчислюються категоріальні групи, які будуть використані як фактори в таблицях.

6.3 Обчислення таблиць

Після того, як всі дані підготовлені, можна розраховувати цільові таблиці.

Основні відомості інвентаризаційних таблиць наведені в таблиці 6. У стовпчику **Table** перераховані номери таблиць результатів. Тут знак /3 означає кожне третє число діапазону. **Значення** - це цільове значення, яке потрібно оцінити. **Об'єкт** - об'єкт збору даних. Стовпці стовпчики **X**, **M** та **Y** - це значення, які потрібно обчислити або оцінюються. У випадку, коли присутній показник **Y**, розраховується відношення **XM/Y**.

Середні характеристики дерев та підрозмірів, крім об'єму, зважуються за кількістю та площею відповідно.

Таблиця	Значення	Об'єкт	X	M
1.(01 - 19), 6.(01 - 03)	Загальна площа	підділянка	площа	-
2.(01 - 17)	Посівні площі	підділянка	об'єм_кількість	-
3.(01-16/3, 13.1)	Дерево	дерево	об'єм	кількість
3.(02-17/3, 14.1), 6.(04, 05)	Кількість дерев	дерево	кількість	-
3.(03-18/3)	Середній об'єм	дерево	об'єм	кількість
5.(01, 03, 10)	Середній запас на 1 га	підділянка	об'єм_кількість	-
5.02	Середній вік деревостану	підділянка	вік деревостану	площа
5.04	Середній діаметр деревостану	підділянка	діаметр деревостану	площа
5.05	Середній діаметр дерев	дерево	діаметр	кількість
5.06	Середня висота деревостану	підділянка	висота деревостану	площа
5.07	Середня висота дерев	дерево	висота	кількість
5.08	Середній бонітет ділянки	підділянка	бонітет ділянки	площа
5.09	Середня густина	підділянка	щільність деревостану	площа
7.01	Загальний сухий об'єм деревостану	підділянка	сухий об'єм_кількість	-
7.02	Середній сухий об'єм деревостану	підділянка	сухий об'єм_кількість	-
7.03	Загальний об'єм пошкоджених дерев	дерево	об'єм	кількість
7.04	Загальний об'єм сухостійних дерев	дерево	об'єм	кількість
7.05	Загальний обсяг сухостою	дерево	об'єм	кількість
8.(01, 02)	Кількість лісовідновлювальних дерев	дерево	кількість	-

Табл. 6: Основні відомості про таблиці інвентаризації

Точне визначення цільових таблиць визначається у файлі outputTables.csv. Значення його стовпців описано у таблиці 7. Для того, щоб скласти нову таблицю, необхідно додати новий рядок у файл outputTables.csv вибравши необхідні параметри.

6.4 Приклади оцінок площ

Нехай таблиця 1.01 "Площа лісових та нелісових земель" представляє розподіл загальної площі регіону за різними категоріями земель. В такому випадку X_{ij} , $j = 1, \dots, N_i$ - площі підділянок у тракті i , а L_{ij} - категорія земель, визначена для ділянки j в тракті i . Для оцінки загальної площі

Назва та Значення	
number	Номер таблиці
main_title	Назва таблиці
dimensions	Назва розмірності характеристики
data_frame	Таблиця полів, що використовується
condition	Назва змінної індикатора для визначення конкретного домену
factor1	Перший фактор для вказівки піддоменів
factor2	Другий фактор для вказівки піддоменів
factor3	Третій фактор для вказівки піддоменів
table_mode	Режим таблиці, один з наступних: - none : обчислює загальну суму або відношення (з двома характеристиками) зна - row : обчислює відсотки для кожного рядка - cv : обчислює відсотки для кожного рядка - cv : обчислити коефіцієнти варіації
характеристика1	Назва характеристики
характеристика2	Назва характеристики для знаменника
multiplier	Додатковий постійний множник
точність	Найвища точність для округлення

Табл. 7: Параметри таблиці інвентаризації

категорії k визначимо умовний показник $I_{ij} = \mathbf{1}_{\{L_{ij}=k\}}$. Тоді формули (7 - 10) дадуть оцінки загальної площі та коефіцієнта варіації з типом k .

Обчислення площ для інших таблиць будуть відрізнятися лише визначенням індикаторів. Наприклад, у таблиці 1.04 розглядається лише тип вкритих лісовою рослинністю земель "Площа дерево-станів домінуючих деревних порід у розрізі користувачів"(тобто код групи_категорій_земель = 11 та відносний показник $I_{ij} = \mathbf{1}_{\{L_{ij}=11\}}$. Таблиця має два фактори: Main_Species та Name_of_User. Таким чином визначимо їх значення через MS_{ij} та NU_{ij} відповідно. Тоді, оцінки для Main_Species m та Name_of_User n можна отримати, використовуючи показники $I_{ij} = \mathbf{1}_{\{L_{ij}=11\}}\mathbf{1}_{\{MS_{ij}=m\}}\mathbf{1}_{\{NU_{ij}=n\}}$ та формули (7 - 10).

6.5 Багаторічні інвентаризації

Кожен п'ятий рік інвентаризація проводиться на одних і тих же ділянках. Таким чином, ми можемо оцінити динаміку лісів за період $t_2 - t_1 = 5$ років. Нехай V_1 та V_2 - об'єми дерев при цих інвентаризації. Тоді ми можемо обчислити такі характеристики:

1. **Приріст.** Для кожного живого дерева, яке було підраховано в момент часу t_1 і дожило до моменту часу t_2 , обчислимо річний приріст як

$$Growth_volume = (V_2 - V_1)/(t_2 - t_1)$$

Для консервативної оцінки ми нехтуємо приростом підросту, який перетинає мінімальний поріг DBH між моментом часу t_1 та моментом часу t_2

$$Growth_volume_I = 0$$

а також приріст на деревах, що загинули

$$Growth_volume_M = 0$$

або були зрубані між моментом часу t_1 та моментом часу t_2

$$Growth_volume_C = 0$$

2. **Об'єм зрубаного дерева.** Об'єм дерева вважається об'ємом вирубки між інвентаризаціями, якщо його діаметр у другій інвентаризації став рівним нулю. Для таких дерев річний об'єм рубки становить

$$Cutting_volume = V_1/(t_2 - t_1)$$

3. **Обсяг загибелі.** Обсяг дерева враховується як обсяг загибелі між інвентаризаціями, якщо при першій інвентаризації воно було живим, а при другій інвентаризації визначено як мертво. Для таких дерев за рік відпад становить

$$Mortality_volume = V_2/(t_2 - t_1)$$

За аналогією з однорічними таблицями інвентаризації основні відомості наведені в таблиці 8.

Table	Value	Object	X	M
4.01	Річний приріст	subplot	Приріст_обсягу_кількість	-
4.02	Приріст за рік	дерево	Приріст_обсягу	кількість
4.03	Обсяг рубок за рік	підділянка	Обсяг_рубок_кількість	-
4.04	Обсяг рубок за рік	дерево	Обсяг_рубок	кількість
4.05	Обсяг загибелі за рік	підгрупа	Обсяг_загибелі_кількість	-
4.06	Обсяг смертності за рік	дерево	Обсяг_загибелі	кількість

Табл. 8: Основні відомості про таблиці інвентаризації приросту

7 Перелік питань для подальшого аналізу

- виправити пропущені та некоректні значення в лісі. Розробити спеціальний додаток для цього.
- вдосконалити моделі висоти та об'єму, використовуючи більш складні формули, напр. <https://academic>
- виправити пропущені значення в коефіцієнтах (пропуски в таблицях).
- покращити обробку пропущених значень

Література

- [1] Методи вибірки, 3-тє видання, 1977, В. Г. Кокран, Вайлі, Нью-Йорк, розділ 6.
- [2] Оптимальний дизайн вибірки для великих лісових інвентаризацій, Андріан Ланц, Швейцарський федеральний технологічний інститут, Цюрих, 2000.
- [3] Таблиці об'ємних сортиментів для деревостанів, що стоять. К. Нікітін, Міністерство лісового господарства Української РСР, Київ, 631 сторінка, 1983.