

Міністерство освіти і науки України  
Житомирський державний університет імені Івана Франка

**Інструктивно-методичні матеріали**  
**до лабораторних занять**  
**з освітньої компоненти**  
**«Математичні методи та інформаційні технології в біології»**  
для підготовки фахівців  
першого (бакалаврського) рівня вищої освіти  
Галузь знань Е Природничі науки, математика та статистика  
Спеціальність Е1 Біологія та біохімія  
Освітньо-професійна програма Біологія

Житомир – 2026

*Рекомендовано до друку вченою радою Житомирського державного університету  
імені Івана Франка*

*(протокол № 8 від 27 березня 2026)*

**Рецензенти:**

**Лариса ШЕВЧУК** – доктор біологічних наук, професор, професор кафедри сталої інфраструктури та гідроекології Державного університету «Житомирська політехніка»;

**Олександр КРАТЮК** – доктор біологічних наук, професор, професор кафедри лісового та садово-паркового господарства Поліського національного університету.

**Юлія ШЕЛЮК** – доктор біологічних наук, професор, професор кафедри ботаніки, біоресурсів та збереження біорізноманіття Житомирського державного університету імені Івана Франка;

**Олександр Гарбар, Діана Гарбар**

Інструктивно-методичні матеріали до лабораторних занять з освітньої компоненти «Математичні методи та інформаційні технології в біології». Житомир: Вид-во ЖДУ ім. І. Франка, 2026. 67 с.

Збірник містить інструктивно-методичні матеріали до лабораторних занять з освітньої компоненти «Математичні методи та інформаційні технології в біології». Призначений для студентів 3 курсу денної форми навчання, що здобувають освіту за ОПП Біологія.

**УДК 502**

© Гарбар О.В., 2026  
© Гарбар Д.А., 2026  
© Житомирський державний  
університет імені Івана  
Франка, 2026

## ЗМІСТ

Вступ	4
Критерії оцінювання занять	5
Тема № 1. Ознайомлення з програмним забезпеченням Past для статистичного аналізу	6
Тема № 2. Описові статистики. Порівняння вибірок. Кореляційний аналіз. Графічне представлення даних.	9
Тема № 3. Дисперсійний аналіз	17
Тема № 4. Статистичні методи редукції та класифікації даних	23
Тема № 5. Методи аналізу $\alpha$ - та $\beta$ -біорізноманіття.	30
Тема № 6. Основи роботи з просторовими даними в Q-GIS. Створення макету карти та підготовка до друку.	35
Тема № 7. Приєднання атрибутів за значенням поля таблиці. Візуалізація атрибутивних даних.	48
Тема № 8. Зональні статистики, екстрагування та статистичний аналіз даних растру для набору точок.	53
Тема № 9. Картування оселища інвазивного виду <i>Opuntia humifusa</i> (Cactaceae) на основі біокліматичних даних та алгоритму MAXENT	58

## ВСТУП

Метою вивчення освітньої компоненти «Математичні методи та інформаційні технології в біології» є ознайомлення студентів з сучасними методами статистичного та картографічного аналізу біологічних даних. Студенти мають засвоїти теоретичні основи біометрії і картографії, оволодіти навичками статистичної обробки даних та навичками аналізу та створення тематичних карт з використанням сучасного програмного забезпечення.

Методичні рекомендації до лабораторних занять з освітньої компоненти «Математичні методи та інформаційні технології в біології» містять основні вказівки щодо виконання лабораторних робіт з метою оволодіння базовими навичками аналізу та картографічного представлення біологічних даних.

Збірник розрахований для здобувачів вищої освіти першого (бакалаврського) рівня освіти третього року навчання за ОПІ Біологія.

## КРИТЕРІЇ ОЦІНЮВАННЯ ЗАНЯТЬ

Оцінювання здобувачів вищої освіти здійснюється відповідно до «Положення про критерії та порядок оцінювання навчальних досягнень здобувачів вищої освіти Житомирського державного університету імені Івана Франка згідно з Європейською кредитною трансферно - накопичувальною системою» [https://zu.edu.ua/offic/ocinjuvannya\\_zvo.pdf](https://zu.edu.ua/offic/ocinjuvannya_zvo.pdf).

На кожному лабораторному занятті студент може отримати такі максимальні бали за наступні види робіт:

<b>Вид заняття</b>	<b>Тестовий контроль знань</b>	<b>Усний контроль знань</b>	<b>Виконання та оформлення роботи</b>
<b>Лабораторне заняття</b>	<b>20</b>	<b>30</b>	<b>50</b>
<b>Індивідуальне заняття</b>			<b>100</b>
<b>МКР</b>	<b>100</b>		

## Лабораторна робота №1

**Тема:** Ознайомлення з програмним забезпеченням PAST для статистичного аналізу.

**Мета:** ознайомитися з особливостями роботи з даними у програмному пакеті для статистичного аналізу PAST.

### Теоретичні питання

1. Основні поняття математичної статистики. Змінні та їх вимірювання. Вимірювальні шкали.
2. Поняття про залежність між змінними. Залежні та незалежні змінні. Основні символи змінних та операцій. Поняття масиву даних.
3. Статистичне спостереження: види, способи, принципи та помилки.
4. Генеральна сукупність та вибірка. Репрезентативність вибірки. Статистична значимість.
5. Статистичні таблиці та графіки.

### Хід роботи

**1. Використовуючи наведену інструкцію, ознайомтесь з інтерфейсом та особливостями роботи з даними у програмному забезпеченні для статистичного аналізу PAST.**

PAST (PAlaeontological STatistics) (<https://past.en.lo4d.com/download>) – це безкоштовне програмне забезпечення для аналізу наукових даних із функціями для обробки даних, побудови графіків, одномірної та багатовимірної статистики, аналізу біорізноманіття, аналізу часових рядів, морфометрії, стратиграфії тощо. На жаль, застосунок не має підтримки української мови, проте це компенсується тим, що він не вимагає встановлення (можна працювати навіть із флеш-накопичувача), швидко запускається і є досить компактним. Програмне забезпечення розробив співробітник Музею природознавства (Natural History Museum) міста Осло (Норвегія) Øyvind Hammer. Попри те, що спочатку застосунок був призначений для палеонтологічних досліджень, він набув широкої популярності у всьому світі серед екологів. На сьогодні використовується версія 2020 року – PAST 4.03, значно розширена та доповнена новими можливостями.

### Електронна таблиця та меню «Правка»

PAST має інтерфейс користувача, схожий на електронну таблицю. Дані вводяться як масив комірок, організованих у рядки (горизонтально) і стовпці (вертикально).

### Введення даних

Щоб ввести дані в комірку, клацніть на комірці мишкою та введіть дані. Клітинками також можна переміщатися за допомогою клавіш зі стрілками.

У комірки можна вводити будь-який текст, але більшість функцій потребують чисел. Кома (,) і десяткова крапка (.) приймаються як десяткові роздільники.

У PAST випадки займають рядки, а змінні – стовпці. Кластерний аналіз завжди групуватиме елементи, тобто рядки. Для перемикання між рядками та стовпцями можна використати операцію транспонування.

### Вибір областей

Більшість операцій у PAST виконуються лише над областю масиву, яку ви вибрали (позначили). Якщо ви спробуєте запустити функцію, але не вибрано жодної області, ви отримаєте повідомлення про помилку.

- Рядок вибирається клацанням миші на підпису рядка (крайній лівий стовпець).
- Стовпець вибирається клацанням миші на ярлику стовпця (верхній рядок).
- Кілька рядків вибираються шляхом вибору мітки першого рядка, а потім клацання міток додаткових рядків, утримуючи клавішу Shift. Зауважте, що ви не можете «перетягнути» кілька рядків – це натомість перемістить перший рядок (див. нижче).
- Кілька стовпців позначаються подібним чином, клацаючи на додаткових стовпцях, утримуючи клавішу Shift.

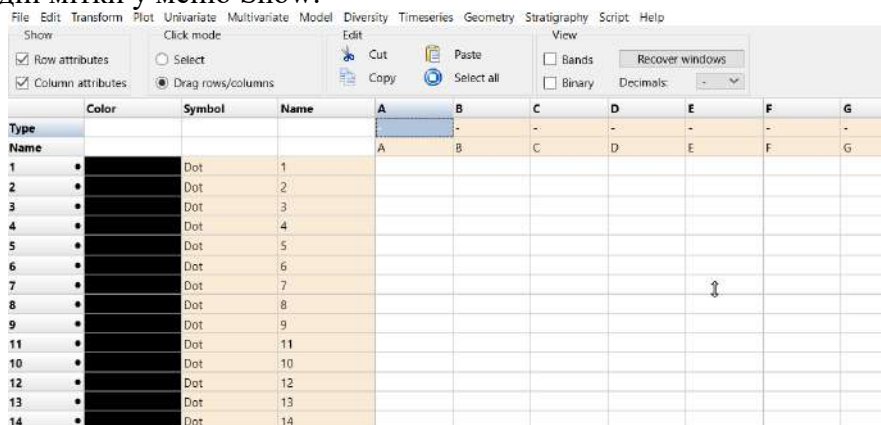
- Весь масив можна виділити, клацнувши верхній лівий кут масиву (порожня сіра клітинка) або вибравши «Виділити все» в меню «Правка».

### Переміщення рядка або стовпця

Рядок або стовпець (включаючи його мітку) можна перемістити, просто клікнувши мітку та перетягнувши її в нове положення в режимі Drag rows/columns (Click mode).

### Перейменування рядків і стовпців

Рядки у PAST нумеруються від 1 до 99, а стовпці позначаються від A до Z. Для правильного підписування графіків слід давати рядкам і стовпцям більш описові, але короткі назви. Назви колонок і рядків, а також тип даних задаються в атрибутах рядків і колонок. Для цього потрібно поставити відповідні мітки у меню Show.



### Збільшення розміру масиву

За замовчуванням PAST має 99 рядків і 26 стовпців. Якщо вам потрібно більше, ви можете додати рядки або стовпці, вибравши «Вставити більше рядків» або «Вставити більше стовпців» у меню «Правка». Рядки/стовпці буде вставлено після позначеної області або внизу/праворуч, якщо область не вибрана. Під час завантаження великих файлів даних рядки та/або стовпці додаються автоматично за потреби.

### Вирізати, скопіювати, вставити

Функції вирізання, копіювання та вставки можна знайти в меню «Правка». Ви можете вирізати/скопіювати дані з електронної таблиці PAST і вставити в інші програми, наприклад Word і Excel. Подібним чином дані з інших програм можна вставити в PAST – вони мають бути у текстовому форматі, розділеному табуляцією.

Усі модулі, що видають графіки, мають кнопку «Копіювати графік». Це помістить графічне зображення в буфер для вставлення в інші програми для редагування зображення. Графіки копіюються за допомогою «розширеного формату метафайлу» у Windows.

### Видалити

Функція видалення (меню «Правка») дозволяє видалити вибрані рядки або стовпці з електронної таблиці. Видалена область не копіюється в буфер вставки.

### Групування (розфарбовування) рядів

Вибрані рядки (випадки) можна позначити кольором в атрибутах рядків. Кожна група також пов'язана з символом (крапка, хрест, квадрат, ромб, плюс, коло, трикутник, лінія, смуга, зафарбований квадрат, зірка, овал, зафарбований трикутник, перевернутий трикутник, зафарбований перевернутий трикутник, зафарбований ромб). Це корисно для відображення різних груп даних на графіках, а також потрібно для багатьох методів аналізу.

Важливо: для методів, які вимагають групування рядків за допомогою кольорів, рядки, що належать до однієї групи, мають бути послідовними.

•	Black	Dot	4	Автономна Ре
•	Black	Dot	5	Автономна Ре
■	Yellow	Square	6	Автономна Ре
•	Turquoise	Dot	7	Автономна Ре
•	Violet	Dot	8	Автономна Ре
•	Wheat	Dot	9	Автономна Ре
•	White	Dot	10	Автономна Ре
•	Whitesmo	Dot	11	Вінницька
•	Yellow	Dot	12	Вінницька
•	Yellowgrec	Dot	13	Вінницька
•	Null	Dot	14	Вінницька
•	Black	Dot	15	Вінницька
•	Black	Dot	16	Вінницька
•	Black	Dot	17	Волинська
•	Black	Dot	18	Волинська
•	Black	Dot	19	Волинська

### Вибір типів даних для стовпців

Вибрані стовпці можна позначити типом даних (безперервний/невизначений, порядковий, номінальний або двійковий) в атрибутах колонок. Це потрібно, лише якщо ви бажаєте використовувати змішані міри подібності/відстані.

### Видалити неінформативні рядки/стовпці

Рядки або стовпці можуть бути неінформативними, особливо щодо багатовимірного аналізу. Такі рядки та стовпці слід видалити. Можна шукати та видаляти: рядки чи стовпці лише з нулями, рядки чи стовпці лише з відсутніми даними («?»), а також рядки чи стовпці лише з однією ненульовою коміркою (одиначні).

### Транспонувати

Функція «Транспонування» в меню «Правка» поміняє місцями рядки та стовпці. Це використовується для перемикання між режимами R і Q в кластерному аналізі, аналізі головних компонент та ін.

### Завантаження та збереження даних

Функція «Відкрити» знаходиться в меню «Файл». Ви також можете перетягнути файл із робочого столу у вікно PAST.

## 2. Використовуючи дані з таблиці 1, створіть базу даних для аналізу у PAST

### Рекомендована література

1. Серікова О.М. Методи обробки статистичних даних : курс лекцій. – Х.: НУЦЗУ, 2019. – 198 с.
2. Янковська Л.В. Математичні методи в екології: навчальний посібник для вузів – Тернопіль: ТНПУ, 2017. – 114с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Тарасова В.В., Парфенцева Н.О., Ковалевська І.М. Екологічна статистика. Підручник. – К.: Центр навчальної літератури, 2013. – 295 с.

## Лабораторна робота № 2

**Тема: Описові статистики. Порівняння вибірок. Кореляційний аналіз. Графічне представлення даних.**

**Мета:** навчитися розраховувати основні описові статистики для однієї чи кількох вибірок однофакторних даних.

### Теоретичні питання

1. Міри центральної тенденції та мінливості.
2. Мода. Медіана та її обчислення.
3. Середнє арифметичне: обчислення та властивості.
4. Інтерпретація мір центральної тенденції. Вибір міри центральної тенденції. Призначення мір мінливості.
5. Розмах та коефіцієнт осциляції.
6. Дисперсія: обчислення та властивості.
7. Стандартне відхилення та коефіцієнт варіації.
8. Аналіз зв'язку між змінними.

### Хід роботи

#### 1. Створіть базу даних і проведіть їх аналіз.

Перенесіть таблицю із даними про середньомісячні температури для основних адміністративних одиниць України у книгу Excell і збережіть її. Для коректної подальшої роботи важливо вставляти дані починаючи з першої клітинки. Порожніх рядків вище заголовку таблиці бути не повинно! Корисно також закріпити рядок із заголовками таблиці, щоб він не зникав при прогортанні таблиці донизу. Для цього зайдіть у меню «Вигляд» - «Закріпити області» - «Закріпити перший рядок».

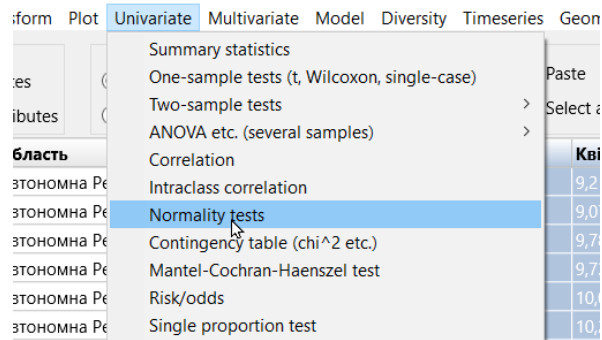
№	Область	Район	Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень	Листопад	Грудень
1	Автономна Республіка Крим	Бахчисарайський	-0,14	0,43	3,51	9,20	14,14	18,38	21,08	20,50	15,96	10,66	5,91	2,20
2	Автономна Республіка Крим	Білогірський	-0,18	1,21	3,50	9,07	13,90	18,63	21,46	21,44	16,76	11,61	6,38	3,04
3	Автономна Республіка Крим	Джанкойський	-1,21	-0,24	3,21	9,78	15,47	19,92	22,41	21,71	16,74	10,99	5,98	2,13
4	Автономна Республіка Крим	Євпаторійський	0,42	1,08	4,24	9,73	15,10	19,86	22,65	21,94	17,26	11,63	7,04	3,40
5	Автономна Республіка Крим	Керченський	0,30	0,90	3,77	10,06	15,38	20,26	23,07	22,54	17,59	11,88	6,47	2,95
6	Автономна Республіка Крим	Курманський	1,17	1,92	4,60	10,24	15,40	20,00	22,91	22,47	17,79	12,07	7,51	4,12
7	Автономна Республіка Крим	Перекопський	-0,74	0,13	3,68	9,97	15,49	19,99	22,50	21,66	16,86	10,92	6,09	2,25
8	Автономна Республіка Крим	Сімферопольський	1,08	1,50	4,03	9,42	14,27	18,72	21,67	21,25	16,73	11,40	6,71	3,36
9	Автономна Республіка Крим	Феодосійський	0,01	1,30	3,85	9,71	14,77	19,58	22,38	22,20	17,40	12,03	6,61	3,22
10	Автономна Республіка Крим	Ялтинський	-0,49	-0,24	2,26	7,88	12,50	16,62	19,34	18,97	14,60	9,75	4,86	1,51
11	Вінницька	Вінницький	-5,59	-3,87	0,82	7,81	13,97	17,01	18,14	17,67	13,48	7,44	1,91	-2,61
12	Вінницька	Гайсинський	-4,85	-4,01	0,80	7,81	13,99	17,08	18,27	17,50	13,45	7,35	2,79	-2,09
13	Вінницька	Жмеринський	-4,74	-3,51	0,91	7,99	14,14	16,93	18,19	17,69	13,27	7,59	1,93	-2,18
14	Вінницька	Могилів-Подільський	-3,87	-2,64	1,75	8,71	14,58	17,60	18,93	18,32	13,98	8,28	2,59	-1,49
15	Вінницька	Тульчинський	-4,47	-3,28	1,04	8,19	13,98	17,28	18,48	17,78	13,73	7,83	2,39	-1,85
16	Вінницька	Хмільницький	-5,18	-3,96	0,65	7,79	14,08	16,90	17,98	17,59	13,11	7,27	1,41	-2,62
17	Волинська	Володимир-Волинський	-3,64	-2,82	1,35	7,50	13,38	16,31	17,63	17,07	12,74	7,81	2,51	-1,43
18	Волинська	Камінь-Каширський	-3,88	-2,91	0,79	7,40	13,29	16,72	18,01	17,87	13,19	8,96	3,29	-1,64
19	Волинська	Ковельський	-3,92	-2,95	1,22	7,28	13,34	16,56	17,50	17,19	12,60	8,07	2,84	-1,47
20	Волинська	Луцький	-3,82	-2,96	0,92	7,60	13,28	16,52	17,92	17,69	13,29	8,71	3,09	-1,47
21	Дніпропетровська	Дніпровський	-4,36	-3,45	1,40	9,69	16,01	19,62	21,29	20,54	15,30	8,51	2,01	-2,20
22	Дніпропетровська	Кам'янський	-4,83	-3,81	0,94	9,11	15,32	18,84	20,37	19,67	14,73	8,12	1,91	-2,36
23	Дніпропетровська	Криворізький	-3,94	-2,92	1,66	9,58	15,71	19,33	21,11	20,38	15,43	8,87	2,60	-1,61
24	Дніпропетровська	Нікопольський	-3,62	-2,53	2,04	9,91	16,13	19,94	21,82	21,23	16,20	9,43	3,31	-0,96
25	Дніпропетровська	Новомосковський	-4,92	-4,08	1,00	9,45	15,74	19,11	20,78	20,15	14,78	8,07	1,45	-2,90
26	Дніпропетровська	Павлоградський	-5,10	-4,23	0,93	9,35	15,48	18,91	20,77	20,08	14,73	8,05	1,53	-2,89
27	Дніпропетровська	Синельниківський	-5,06	-3,99	1,07	9,37	15,34	18,92	20,95	20,27	15,01	8,40	1,88	-2,66
28	Донецька	Бахмутський	-5,52	-4,90	0,61	9,63	15,54	19,30	20,88	19,73	14,43	7,64	0,98	-3,15

29	Донецька	Волноваський	-5,19	-3,89	0,95	9,34	15,25	19,16	21,48	21,12	15,74	9,18	2,54	-2,15
30	Донецька	Горлівський	-5,75	-4,60	0,55	9,41	15,24	19,31	21,28	21,02	15,49	8,71	1,62	-2,75
31	Донецька	Донецький	-5,18	-3,87	1,09	9,65	15,44	19,50	21,56	21,61	16,04	9,34	2,11	-2,10
32	Донецька	Кальміуський	-4,49	-3,38	1,33	9,80	15,73	19,90	22,07	21,97	16,43	9,67	2,52	-1,56
33	Донецька	Краматорський	-5,39	-4,65	0,67	9,47	15,39	19,06	20,74	19,55	14,22	7,60	1,08	-3,09
34	Донецька	Маріупольський	-4,08	-3,12	1,43	9,53	15,64	19,87	22,25	21,62	16,27	9,60	3,03	-1,37
35	Донецька	Покровський	-5,38	-4,29	0,82	9,41	15,24	18,93	20,89	20,12	14,82	8,22	1,62	-2,78
36	Житомирська	Бердичівський	-5,33	-3,93	0,64	7,64	13,88	16,89	17,89	17,55	13,12	7,35	1,63	-2,61
37	Житомирська	Житомирський	-5,28	-4,02	0,44	7,37	13,86	16,70	17,76	17,34	12,57	7,18	1,74	-2,46
38	Житомирська	Коростенський	-6,03	-4,55	-0,01	7,01	13,58	16,38	17,31	16,53	11,85	6,83	1,96	-2,75
39	Житомирська	Новоград-Волинський	-5,40	-4,37	0,26	6,57	13,31	15,93	17,09	16,87	11,68	6,97	1,76	-2,63
40	Закарпатська	Берегівський	-1,97	-0,12	4,87	10,13	15,30	18,24	19,91	19,52	15,41	10,27	4,32	-0,16
41	Закарпатська	Мукачівський	-3,03	-1,76	2,69	7,56	12,73	15,76	17,31	16,95	13,21	8,66	3,21	-1,36
42	Закарпатська	Рахівський	-5,06	-4,02	-0,85	4,32	9,67	12,67	14,24	14,11	10,45	6,42	0,87	-3,66
43	Закарпатська	Тячівський	-4,31	-3,23	0,58	5,65	10,91	13,94	15,48	15,22	11,51	7,22	1,61	-2,82
44	Закарпатська	Ужгородський	-3,05	-1,74	2,67	7,54	12,70	15,57	17,10	16,71	12,90	8,38	2,96	-1,47
45	Закарпатська	Хустський	-3,82	-2,71	1,36	6,30	11,47	14,60	16,17	15,88	12,19	7,84	2,33	-2,16
46	Запорізька	Бердянський	-3,75	-2,43	1,76	9,37	15,34	19,74	22,06	21,30	16,29	9,70	3,95	-0,63
47	Запорізька	Василівський	-3,72	-2,66	1,99	9,61	15,68	19,54	21,55	20,83	15,80	9,26	3,39	-0,96
48	Запорізька	Запорізький	-4,45	-3,39	1,52	9,62	15,52	19,28	21,35	20,70	15,48	8,86	2,22	-2,11
49	Запорізька	Мелітопольський	-3,39	-2,41	2,04	9,23	15,46	19,82	21,88	21,14	15,88	9,48	4,49	-0,23
50	Запорізька	Пологівський	-4,81	-3,43	1,37	9,19	14,86	18,88	21,35	20,90	15,75	9,29	3,06	-1,72
51	Івано-Франківська	Верховинський	-6,18	-5,08	-2,61	2,83	8,29	11,43	13,06	13,15	9,63	5,88	0,26	-4,53
52	Івано-Франківська	Івано-Франківський	-4,25	-2,73	1,20	7,28	12,76	15,69	17,12	16,81	12,94	8,03	2,40	-1,91
53	Івано-Франківська	Калузький	-4,33	-3,32	0,15	5,60	10,74	13,99	15,56	15,36	11,66	7,37	1,82	-2,50
54	Івано-Франківська	Коломийський	-4,73	-2,58	1,02	7,61	12,97	16,16	17,76	17,68	13,85	9,13	2,96	-2,28
55	Івано-Франківська	Косівський	-4,93	-2,96	-0,18	6,07	11,26	14,56	16,23	16,34	12,75	8,63	2,52	-2,84
56	Івано-Франківська	Надвірнянський	-5,08	-3,98	-1,09	4,44	9,77	12,91	14,50	14,48	10,88	6,84	1,19	-3,42
57	Київська	Білоцерківський	-5,33	-3,68	0,88	8,00	13,86	17,80	19,09	19,19	14,44	9,16	3,36	-1,99
58	Київська	Бориспільський	-4,94	-4,00	0,69	8,47	14,88	17,82	19,18	18,59	13,63	7,55	1,33	-2,65
59	Київська	Броварський	-4,87	-4,05	0,74	8,43	14,76	17,68	19,02	18,35	13,34	7,35	1,08	-2,81
60	Київська	Бучанський	-5,65	-3,86	0,42	7,73	14,40	17,19	18,34	17,78	12,82	7,43	2,17	-2,27
61	Київська	Вишгородський	-5,21	-3,83	0,57	7,81	14,52	17,26	18,55	17,72	12,75	7,14	1,83	-2,40
62	Київська	Київська	-4,27	-3,23	1,25	8,78	15,00	18,12	19,36	18,78	13,73	7,98	1,73	-2,03
63	Київська	Обухівський	-4,80	-3,66	0,88	8,50	14,66	17,85	19,10	18,72	13,91	8,08	1,89	-2,35
64	Київська	Фастівський	-5,58	-3,78	0,55	7,95	14,10	17,47	18,46	18,47	13,63	8,34	2,55	-2,17
65	Київська	Чорнобильська зона відчуження	-5,24	-4,02	0,59	7,64	14,27	17,10	18,51	17,45	12,61	6,87	1,66	-2,68
66	Кіровоградська	Голованівський	-4,63	-3,65	1,35	8,71	14,58	18,06	19,42	18,97	14,22	8,35	2,76	-1,76
67	Кіровоградська	Кропивницький	-4,55	-3,77	1,02	8,87	15,02	18,51	20,11	19,59	14,60	8,27	1,89	-2,32
68	Кіровоградська	Новоукраїнський	-5,00	-4,30	0,83	8,27	14,39	18,35	19,68	19,59	14,40	8,66	2,73	-1,95
69	Кіровоградська	Олександрійський	-4,60	-3,63	1,06	9,13	15,41	18,95	20,72	19,97	14,77	8,30	2,05	-2,34
70	Луганська	Алчевський	-5,86	-5,11	0,28	9,39	15,24	19,26	21,11	20,26	14,75	7,95	1,21	-3,07
71	Луганська	Довжанський	-6,10	-5,22	-0,04	9,12	15,08	19,34	21,50	21,33	15,73	9,15	2,06	-2,82
72	Луганська	Луганський	-5,84	-5,02	0,22	9,36	15,15	19,37	21,33	20,71	15,18	8,55	1,65	-2,86
73	Луганська	Ровеньківський	-6,14	-5,00	0,12	9,14	14,98	19,24	21,34	21,29	15,71	9,17	2,05	-2,69
74	Луганська	Сватівський	-6,60	-6,08	-0,29	9,11	15,26	18,78	20,32	19,14	13,60	6,79	0,45	-4,00
75	Луганська	Северодонецький	-5,83	-5,29	0,30	9,48	15,44	19,22	20,91	19,66	14,20	7,41	0,92	-3,33
76	Луганська	Старобільський	-6,70	-6,03	-0,39	9,34	15,54	19,28	20,92	19,75	14,16	7,07	0,61	-3,81
77	Луганська	Щастинський	-6,02	-5,33	0,11	9,49	15,42	19,38	21,11	19,99	14,52	7,65	1,04	-3,28
78	Львівська	Дрогобицький	-3,89	-2,63	0,78	6,34	10,99	14,75	16,47	16,47	12,87	8,74	3,15	-1,67
79	Львівська	Золочівський	-4,40	-3,35	0,92	7,18	12,92	15,61	16,68	16,29	12,37	7,46	2,89	-1,39
80	Львівська	Львівський	-3,89	-2,69	1,40	7,30	12,94	15,72	17,14	16,70	12,71	7,77	2,70	-1,45
81	Львівська	Самбірський	-3,70	-2,81	1,10	6,04	11,21	14,45	16,00	15,69	12,17	7,94	2,77	-1,78
82	Львівська	Стрийський	-4,18	-3,01	0,80	6,38	11,40	14,72	16,27	16,08	12,34	7,92	2,45	-2,02

83	Львівська	Червоноградський	-3,76	-2,85	1,42	7,40	13,32	16,03	17,40	16,78	12,66	7,64	2,82	-1,29
84	Львівська	Яворівський	-3,53	-2,22	1,68	7,18	12,80	15,73	17,36	16,93	13,04	8,27	3,15	-1,28
85	Миколаївська	Баштанський	-3,39	-2,58	1,97	9,40	15,43	19,17	21,27	20,70	15,50	9,13	2,76	-1,30
86	Миколаївська	Вознесенський	-3,36	-2,54	2,35	9,53	15,46	19,39	21,30	21,12	15,85	9,85	3,32	-0,75
87	Миколаївська	Миколаївський	-2,61	-1,69	2,76	9,97	16,00	19,97	22,12	21,70	16,46	10,06	3,72	-0,32
88	Миколаївська	Первомайський	-3,80	-2,81	2,05	9,20	14,88	18,90	20,63	20,56	15,36	9,57	3,04	-1,12
89	Одеська	Березівський	-3,30	-2,13	2,31	9,48	15,49	19,17	20,91	20,83	15,86	9,76	3,61	-0,65
90	Одеська	Білгород-Дністровський	-1,64	-0,46	3,09	9,38	15,30	19,63	21,22	20,56	16,44	10,76	5,58	1,02
91	Одеська	Болградський	-1,81	-0,55	3,41	9,94	15,82	19,61	21,23	20,75	16,42	10,63	5,10	0,58
92	Одеська	Ізмаїльський	-0,68	0,46	4,18	10,46	16,22	20,20	21,90	21,31	17,07	11,44	5,79	1,43
93	Одеська	Одеський	-2,32	-1,20	2,73	9,29	15,19	19,48	21,36	20,89	16,34	10,45	4,63	0,31
94	Одеська	Подільський	-3,93	-3,02	1,60	8,97	14,56	17,89	19,22	18,92	14,50	8,51	3,01	-1,46
95	Одеська	Роздільнянський	-3,67	-2,30	2,01	9,33	15,27	18,86	20,16	20,23	15,53	9,55	3,87	-0,75
96	Полтавська	Кременчуцький	-4,73	-3,82	1,05	9,18	15,55	18,93	20,56	19,84	14,66	8,15	1,96	-2,40
97	Полтавська	Лубенський	-5,31	-4,40	0,66	8,88	15,26	18,43	19,79	18,95	13,87	7,51	1,16	-2,96
98	Полтавська	Миргородський	-5,71	-4,79	0,35	8,64	15,06	18,13	19,50	18,66	13,59	7,19	0,85	-3,38
99	Полтавська	Полтавський	-5,57	-4,55	0,50	8,85	15,31	18,41	19,82	19,07	13,90	7,37	1,15	-3,26
100	Рівненська	Вараський	-3,86	-2,89	0,85	7,67	13,43	16,68	18,25	17,85	13,22	8,65	2,70	-1,91
101	Рівненська	Дубенський	-3,83	-2,96	0,93	7,71	13,16	16,57	17,83	17,89	13,70	9,11	3,24	-1,35
102	Рівненська	Рівненський	-4,00	-3,19	1,02	7,68	13,60	16,56	17,82	17,43	12,87	7,81	1,83	-2,00
103	Рівненська	Сарненський	-4,48	-3,46	0,81	7,44	13,35	16,42	17,63	17,04	12,41	7,55	1,94	-2,24
104	Севастопільська	Севастопільська	0,56	1,29	4,51	10,13	15,43	19,89	22,60	21,90	17,23	11,68	7,04	3,31
105	Сумська	Конотопський	-6,32	-5,56	-0,55	7,89	14,40	17,62	18,73	17,71	12,62	6,45	0,09	-4,02
106	Сумська	Охтирський	-6,34	-5,32	-0,33	8,11	14,57	17,75	19,04	18,01	12,78	6,56	0,28	-4,18
107	Сумська	Роменський	-6,59	-5,61	-0,45	8,11	14,72	17,77	18,94	18,08	13,03	6,73	0,52	-3,81
108	Сумська	Сумський	-6,37	-5,57	-0,58	7,97	14,46	17,76	18,95	17,91	12,70	6,45	-0,14	-4,32
109	Сумська	Шосткинський	-6,87	-6,06	-1,01	7,17	13,95	17,19	18,36	17,17	11,98	5,95	-0,04	-4,49
110	Тернопільська	Кременецький	-3,99	-3,04	0,76	7,59	12,98	16,49	17,72	17,99	13,86	9,23	2,76	-1,83
111	Тернопільська	Тернопільський	-4,71	-3,44	0,65	7,20	13,00	15,78	16,91	16,76	12,79	7,72	2,02	-2,22
112	Тернопільська	Чортківський	-4,53	-2,77	1,21	7,83	13,45	16,46	17,84	17,61	13,63	8,50	2,47	-1,96
113	Харківська	Богодухівський	-6,61	-5,43	-0,40	8,20	14,76	17,90	19,28	18,35	12,99	6,62	0,69	-4,10
114	Харківська	Ізюмський	-5,82	-5,05	0,33	9,15	15,39	18,96	20,47	19,29	13,85	7,23	0,98	-3,34
115	Харківська	Красноградський	-5,65	-4,84	0,37	8,91	15,26	18,52	20,15	19,57	14,09	7,49	1,01	-3,47
116	Харківська	Куп'янський	-7,10	-6,17	-0,48	8,41	15,14	18,43	19,78	18,73	13,19	6,52	0,67	-4,13
117	Харківська	Лозівський	-5,76	-4,91	0,38	9,08	15,36	18,77	20,42	19,64	14,22	7,53	1,18	-3,24
118	Харківська	Харківський	-6,45	-5,50	-0,26	8,83	15,20	18,80	20,29	19,22	13,64	6,98	0,72	-3,61
119	Харківська	Чугувський	-6,86	-5,89	-0,41	8,66	15,30	18,75	20,14	19,06	13,51	6,79	0,94	-3,67
120	Херсонська	Бериславський	-3,14	-2,30	2,15	9,57	15,53	19,41	21,70	20,95	15,79	9,37	3,34	-0,67
121	Херсонська	Генічеський	-2,84	-1,96	2,14	9,43	15,77	20,02	22,11	21,15	16,05	9,71	4,45	0,40
122	Херсонська	Каховський	-2,91	-1,89	2,37	9,64	15,57	19,74	21,93	21,09	16,02	9,66	3,94	0,01
123	Херсонська	Скадовський	-1,24	-0,67	3,24	9,95	15,68	20,05	22,37	21,62	16,74	10,62	4,74	0,90
124	Херсонська	Херсонський	-2,12	-1,33	2,92	9,97	15,80	19,95	22,28	21,46	16,35	10,01	3,91	0,04
125	Хмельницька	Кам'янець-Подільський	-4,61	-2,68	1,39	7,86	13,64	16,73	18,23	17,99	13,94	8,92	3,13	-1,61
126	Хмельницька	Хмельницький	-4,65	-3,53	0,76	7,61	13,69	16,47	17,71	17,47	13,10	7,76	1,71	-2,42
127	Хмельницька	Шепетівський	-4,41	-3,53	0,77	7,58	13,68	16,51	17,74	17,48	12,89	7,60	1,54	-2,32
128	Черкаська	Звенигородський	-5,56	-4,33	0,45	7,96	14,48	17,33	18,75	18,09	13,36	7,50	2,53	-2,45
129	Черкаська	Золотонський	-5,39	-4,40	0,54	8,65	15,14	18,16	19,51	18,82	13,96	7,60	1,71	-2,62
130	Черкаська	Уманський	-4,90	-3,83	1,13	8,23	14,14	17,79	19,32	18,72	13,96	8,36	2,95	-2,05
131	Черкаська	Черкаський	-5,09	-4,32	0,43	8,22	14,63	17,58	18,98	18,33	13,66	7,43	1,86	-2,60
132	Чернівецька	Вижницький	-5,28	-3,49	-0,83	5,25	10,60	13,95	15,55	15,57	11,93	7,79	1,76	-3,27
133	Чернівецька	Дністровський	-3,94	-2,04	1,94	8,39	14,13	17,33	18,79	18,26	14,27	9,16	3,40	-1,12
134	Чернівецька	Чернівецький	-3,85	-2,07	1,85	8,35	13,74	17,05	18,54	18,08	14,03	8,87	2,61	-1,69
135	Чернігівська	Корюківський	-6,08	-5,20	-0,30	7,60	13,90	17,16	18,39	17,31	12,26	6,27	0,07	-3,77
136	Чернігівська	Ніжинський	-5,61	-4,86	0,17	8,04	14,41	17,42	18,70	17,83	12,83	6,75	0,44	-3,43

137	Чернігівська	Новгород-Сіверський	-6,56	-5,84	-0,74	7,18	13,65	16,86	18,13	17,04	11,94	5,96	-0,10	-4,25
138	Чернігівська	Прилуцький	-5,90	-5,12	0,12	8,14	14,59	17,54	18,89	18,03	13,07	6,82	0,56	-3,42
139	Чернігівська	Чернігівський	-5,56	-4,60	0,17	7,80	14,09	17,35	18,62	17,57	12,59	6,66	0,71	-3,24

**Оцінка нормальності розподілу.** Відкрийте файл з даними у PAST і виділіть всі клітинки з даними. Зайдіть у вкладку Univariate і оберіть Normality test.



Отриманий результат свідчить про відхилення від нормального розподілу для усіх змінних. Тому застосовувати параметричні методи до цих даних не рекомендується.

	Січень	Лютий	Березень	Квітень	Травень	Червень
N	139	139	139	139	139	139
Shapiro-Wilk W	0,8998	0,9329	0,9505	0,9136	0,8558	0,9305
p(normal)	3,353E-08	3,507E-06	7,073E-05	2,06E-07	2,481E-10	2,407E-06
Anderson-Darling A	4,106	2,619	2,686	2,596	5,499	2,222
p(normal)	2,907E-10	1,223E-06	8,421E-07	1,393E-06	1,259E-13	1,157E-05
p(Monte Carlo)	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
Lilliefors L	0,1408	0,1185	0,1296	0,1172	0,1244	0,1085
p(normal)	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
p(Monte Carlo)	0,0001	0,0002	0,0001	0,0004	0,0001	0,0009
Jarque-Bera JB	43,33	22,96	13,51	65,69	100,1	24,73
p(normal)	3,904E-10	1,033E-05	0,001164	5,439E-15	1,814E-22	4,273E-06
p(Monte Carlo)	0,0001	0,0013	0,0064	0,0003	0,0001	0,0022

Для розрахунку параметричних статистик оберіть у вкладці Univariate опцію Summary statistics. Ця функція обчислює низку основних описових статистичних даних для однієї чи кількох вибірок однофакторних даних:

- N:** число випадків
- Min:** мінімальне значення
- Max:** максимальне значення
- Sum:** сума
- Mean:** середнє арифметичне
- Std. error:** стандартна похибка середнього.
- Variance:** дисперсія.
- Stand. dev.:** стандартне відхилення.
- Median:** медіана.
- 25 prcntil:** 25 процентіль.
- 75 prcntil:** 75 процентіль.
- Skewness:** асиметрія.
- Kurtosis:** ексцес.
- Geom. mean:** геометричне середнє.
- Coeff.var:** коефіцієнт варіації.

Кожен зразок повинен мати принаймні 3 значення та займати один стовпець у електронній таблиці. Стовпці не обов'язково повинні містити однакову кількість значень.

	Січень	Лютий	Березень	Квітень	Травень	Червень
N	139	139	139	139	139	139
Min	-7,1	-6,17	-2,61	2,83	8,29	11,43
Max	1,17	1,92	4,87	10,46	16,22	20,26
Sum	-597,61	-450,89	161,49	1166,98	1999,82	2475,93
Mean	-4,299353	-3,243813	1,161799	8,39554	14,31525	17,81245
Std. error	0,1489283	0,1483044	0,1090623	0,1091083	0,1217361	0,148052
Variance	3,082968	3,057195	1,653348	1,654742	2,059935	3,046794
Stand. dev	1,755838	1,748484	1,285826	1,286368	1,435247	1,745507
Median	-4,63	-3,46	0,92	8,64	14,66	18,06
25 prcntil	-5,52	-4,37	0,42	7,64	13,65	16,7
75 prcntil	-3,75	-2,58	1,76	9,4	15,38	19,26
Skewness	1,219894	0,9335199	0,675967	-1,226199	-1,580698	-0,9557579
Kurtosis	1,379077	0,8159606	0,8096979	2,466828	2,885709	0,9083257
Geom. mean	0	0	0	8,275481	14,23382	17,72059
Coeff. var	-40,8396	-53,90211	110,6755	15,32204	10,026	9,799366

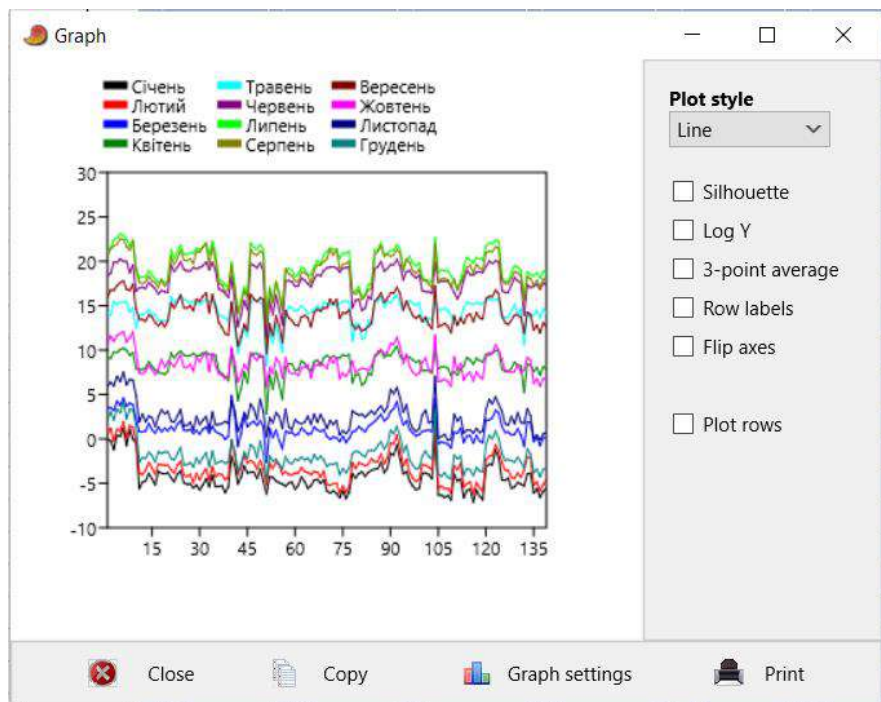
Оскільки для наших даних умова нормальності не виконується, з усіх розрахованих статистик ми можемо використати лише непараметричні показники. Наприклад, медіану.

Якщо оцінити нормальність розподілу по окремих областях, то ми побачимо, що у більшості випадків умова нормальності виконується. Тобто на характер мінливості даних суттєвий вплив має географічний фактор. Тому ми можемо розрахувати і інтерпертувати параметричні статистики для кожної області окремо. Наприклад, описові статистики для АР Крим:

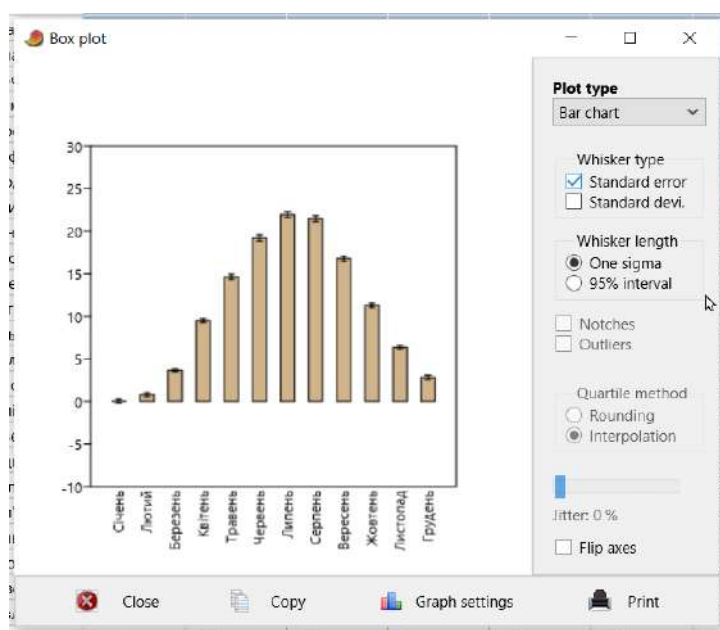
	Січень	Лютий	Березень	Квітень	Травень	Червень	Липень	Серпень	Вересень	Жовтень	Листопад	Грудень
N	8	8	8	8	8	8	8	8	8	8	8	8
Min	-1,21	-0,24	3,21	9,07	13,9	18,38	21,08	20,5	15,96	10,66	5,91	2,13
Max	1,17	1,92	4,6	10,24	15,49	20,26	23,07	22,54	17,79	12,07	7,51	4,12
Sum	0,7	6,93	30,54	77,47	119,15	155,76	177,75	173,51	135,69	91,16	52,09	23,45
Mean	0,0875	0,86625	3,8175	9,68375	14,89375	19,47	22,21875	21,68875	16,96125	11,395	6,51125	2,93125
Std. error	0,2928966	0,2547123	0,1593261	0,1478771	0,237832	0,2667998	0,2562674	0,234021	0,2038508	0,175367	0,1962455	0,2489151
Variance	0,6663071	0,5190268	0,2030796	0,1749411	0,4525125	0,5694571	0,5253839	0,4381268	0,3324411	0,2460286	0,3080982	0,4956696
Stand. dev	0,8264366	0,7204351	0,4506424	0,4182596	0,6726905	0,7546238	0,7248337	0,6619115	0,576577	0,4960127	0,555066	0,7040381
Median	0,08	0,99	3,725	9,755	15,24	19,89	22,455	21,685	16,81	11,505	6,425	2,995
25 prcntil	-0,6	0,205	3,5025	9,255	14,1725	18,6525	21,5125	21,2975	16,7325	10,9375	6,0075	2,2125
75 prcntil	0,915	1,4275	4,1875	10,0375	15,4525	19,9975	22,845	22,3375	17,5075	11,8175	6,9575	3,39
Skewness	-0,1639547	-0,1838419	0,5747719	-0,2939808	-0,6229034	-0,6101962	-0,5112589	-0,4401849	-0,1905057	-0,196855	0,7880836	0,3500735
Kurtosis	-0,7737759	-0,7788718	-0,2070976	-1,269391	-1,870324	-1,906281	-1,298806	0,3501761	0,2125617	-1,28754	-0,1436549	-0,7238008
Geom. mean	0	0	3,794786	9,675788	14,88026	19,45704	22,20831	21,67985	16,95264	11,3855	6,491137	2,857916
Coeff. var	946,7846	83,16712	11,80465	4,31919	4,516596	3,875829	3,262261	3,051865	3,399378	4,352898	8,524722	24,01836

**Розрахуйте самостійно описові статистики середньомісячних температур для всіх областей України.**

Для графічної візуалізації статистичних закономірностей в масиві даних можна побудувати різноманітні графіки. Наприклад, лінійні графіки. Виділіть всі числові дані і в модулі Plot оберіть Graf. Отриманий графік демонструє мінливість середньомісячних температур на території України. По осі x відкладено номери випадків, а по осі y – середньомісячні температури.

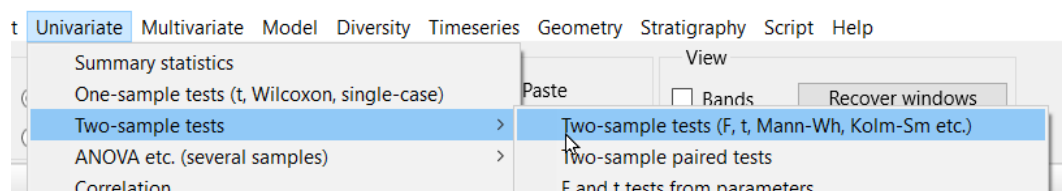


Кращу візуалізацію можна отримати, використавши стовпчасту діаграму (Bar chart).

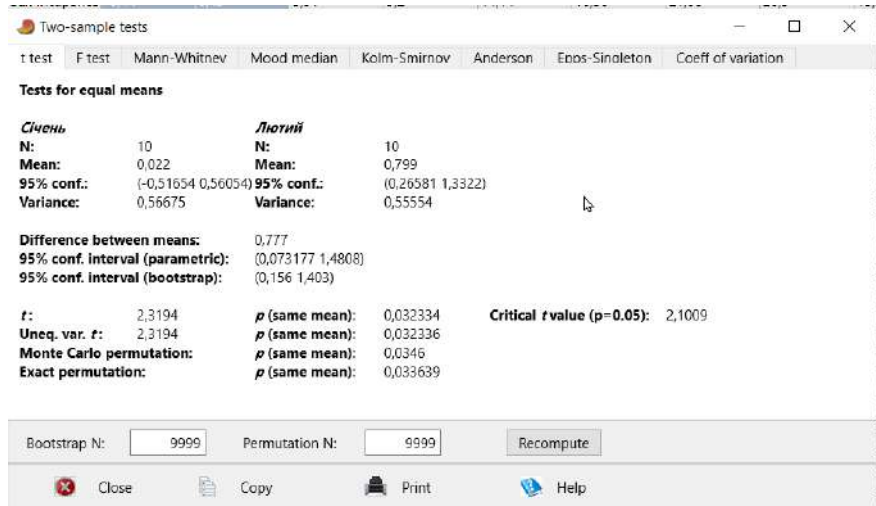


У статистичному аналізі часто доводиться відповідати на питання, чи спостерігаються достовірні відмінності між змінними. Наприклад, чи відрізняються середньомісячні температури січня і лютого в АР Крим? З цією метою використовується тест Стюдента або t-test.

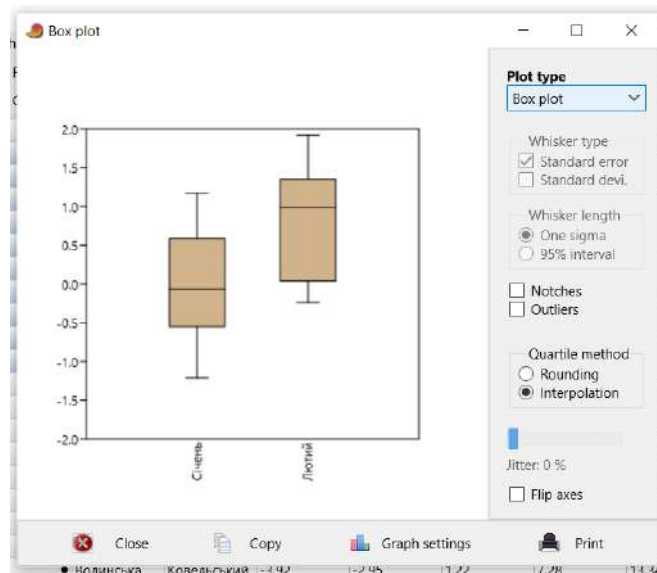
Виділіть необхідну для аналізу частину даних і використайте парний тест Стюдента.



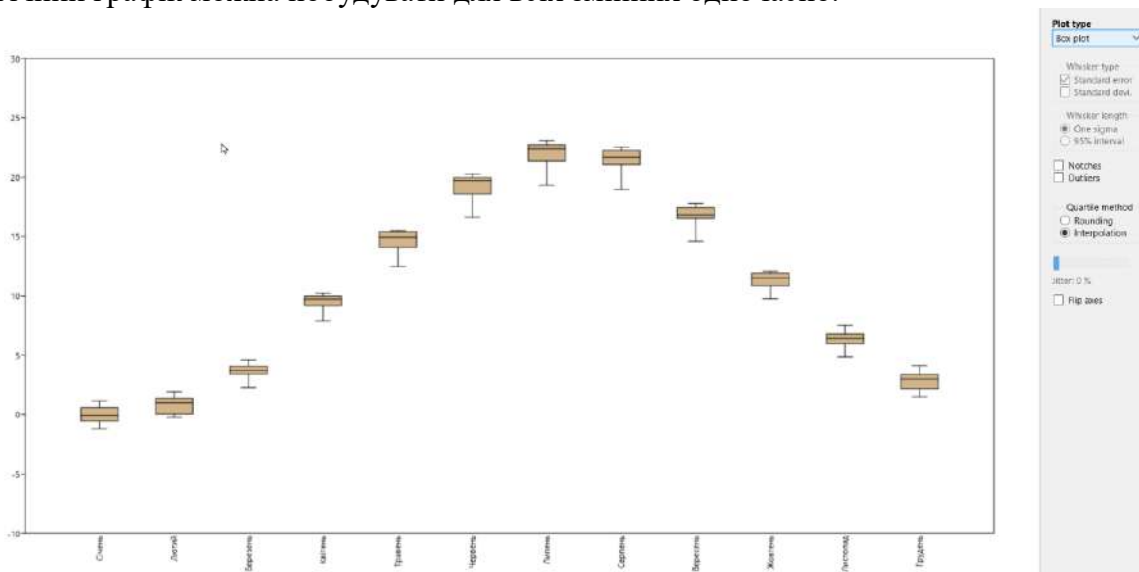
Отримані результати свідчать про наявність достовірних відмінностей середніх температур січня і лютого в АР Крим ( $t=2,32$ ;  $p=0,03$ ).



Ця статистична закономірність добре візуалізується у вигляді коробчастих діаграм. У вкладці Plot оберіть Barchart/Boxplot.

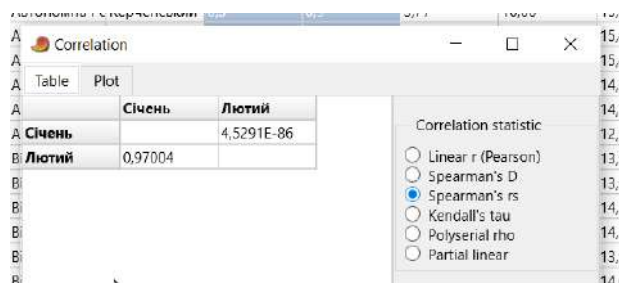


Аналогічний графік можна побудувати для всіх змінних одночасно.



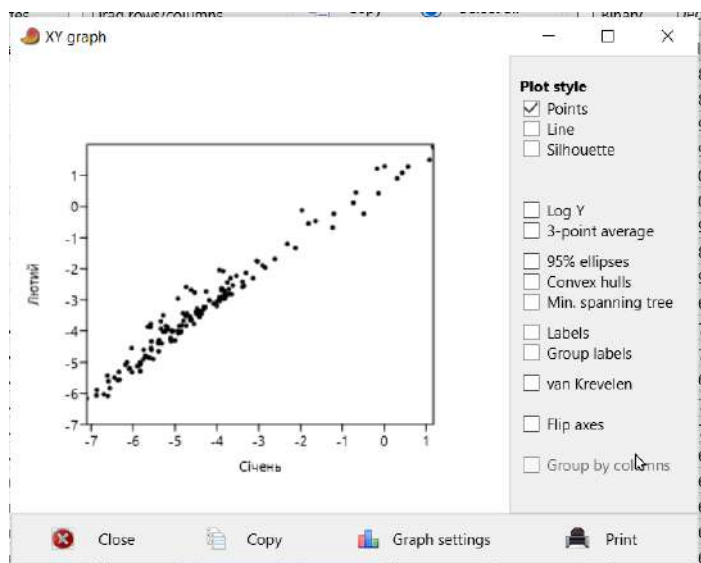
Ми також можемо оцінити наявність зв'язку між змінними, використовуючи модуль кореляційного аналізу. Наприклад, нас цікавить питання, чи є статистичний зв'язок між середніми температурами січня та лютого (або будь-яких інших місяців на ваш розсуд).

Для цього потрібно виділити необхідні для аналізу дані та в модулі Univariate обрати Corellation. Оскільки розподіл даних відхиляється від нормального, потрібно використати непараметричний коефіцієнт кореляції Спірмена. В отриманій таблиці над діагоналлю знаходяться значення рівня достовірності, а під діагоналлю значення обраного коефіцієнта кореляції.

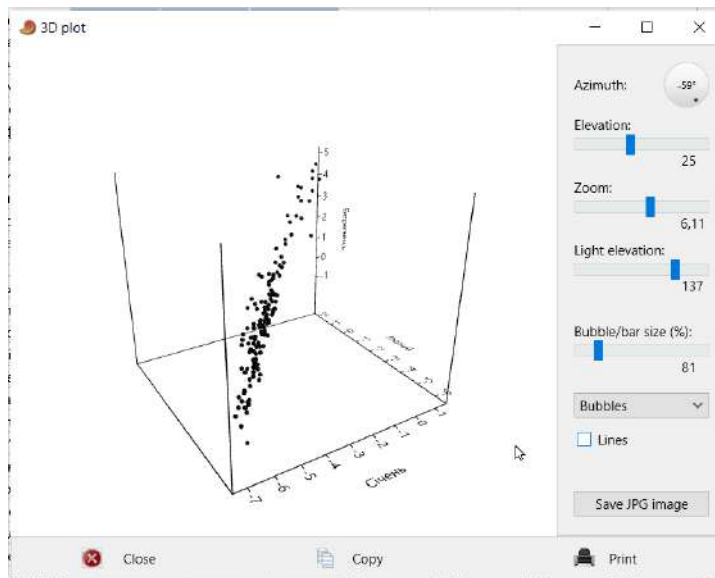


Отриманий результат свідчить про наявність значущого прямопропорційного зв'язку між обраними змінними ( $p < 0,05$ ).

Для візуалізації статистичного зв'язку використовуються ху-графіки. Оберіть необхідну опцію у вкладці Plot. Кожна точка на графіку відображає окремий випадок з масиву даних в координатах двох використаних для аналізу змінних. Чим більший кут нахилу хмари точок – тим сильніший зв'язок.



Ми можемо побудувати також 3Д графіки, використавши три змінні.



**Розрахуйте коефіцієнти кореляції та побудуйте графіки для інших пар змінних (можна виділити одразу весь масив даних і розрахувати всі можливі попарні кореляції).**

### Рекомендована література

1. Серікова О.М. Методи обробки статистичних даних : курс лекцій. – Х.: НУЦЗУ, 2019. – 198 с.
2. Янковська Л.В. Математичні методи в екології: навчальний посібник для вузів – Тернопіль: ТНПУ, 2017. – 114с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Тарасова В.В., Парфенцева Н.О., Ковалевська І.М Екологічна статистика. Підручник. – К.: Центр навчальної літератури, 2013. – 295 с.
5. PAST: Paleontological Statistics Software Package for Education and Data Analysis. [https://palaeo-electronica.org/2001\\_1/past/past.pdf](https://palaeo-electronica.org/2001_1/past/past.pdf)

### Лабораторна робота № 3

**Тема: Дисперсійний аналіз.**

**Мета:** навчитися застосовувати алгоритми дисперсійного аналізу для з'ясування закономірностей мінливості екологічних даних.

#### Теоретичні питання

1. Поняття про дисперсійний аналіз.
2. Однофакторний аналіз.
3. Двофакторний аналіз.

#### Хід роботи

На сьогодні дисперсійний аналіз є одним із основних інструментів для аналізу зв'язків багатовимірних даних. Зокрема, у синекологічних дослідженнях виникає необхідність виявити достовірні відмінності між кількома вибірками, що дозволяє продемонструвати закономірності розподілу видів у просторі (градієнті місцезнаходження) та в часі, відмінності у видовому складі, чисельності, різноманітті, морфології, а також допомагає організувати правильні підходи до охорони рідкісних видів і цінних біотопів тощо.

**Короткий опис можливостей дисперсійного аналізу в синекологічних дослідженнях.** В англійській літературі дисперсійний аналіз зазвичай називають аналізом варіації (Analysis of variance). Його суть полягає в аналізі відхилень усіх одиниць досліджуваної сукупності від середнього арифметичного. Як міра відхилень використовується дисперсія або середній квадрат відхилень. Відхилення, спричинені впливом досліджуваного фактора, порівнюються з величиною відхилень, спричинених випадковими факторами (тобто тими, які дослідник не врахував, але які також можуть впливати на результати). Якщо відхилення, спричинені досліджуваним фактором є більш значущими, ніж випадкові відхилення, вважається, що цей фактор має істотний вплив.

Якщо розглядається більше двох незалежних вибірок, зазвичай використовується однофакторний дисперсійний аналіз (one-way ANOVA) у разі відповідності даних нормальному розподілу або його непараметричні аналоги (тести Краскела-Уолліса, Фрідмана тощо) – в іншому випадку. При цьому нульова гіпотеза стверджує, що між вибірками немає відмінностей, тоді як альтернативна – що відмінності існують хоча б між двома вибірками.

Часто виникає питання: чому аналіз називається однофакторним, якщо порівнюється більше двох вибірок? Якщо є лише одна залежна змінна, наприклад, кількість видів у більш ніж двох біотопах, то маємо справу з однофакторним дисперсійним аналізом (one-way ANOVA). У цьому випадку залежна змінна – це кількість видів, а кілька незалежних – це типи біотопів, у яких вони зафіксовані (наприклад, три).

Для дослідження впливу двох незалежних змінних на одну залежну застосовується двофакторний дисперсійний аналіз (two-way ANOVA). Наприклад, якщо змінними є температура та вологість, то в однофакторному аналізі оцінюється їхній вплив на залежну змінну окремо, тоді як у двофакторному враховується ще й їхній спільний вплив. Багатовимірний дисперсійний аналіз (multivariate analysis of variance, MANOVA) застосовується у випадках, коли є більше ніж одна залежна змінна.

**Умови застосування параметричного однофакторного дисперсійного аналізу (one-way ANOVA).** Перша необхідна умова – відповідність даних закону нормального розподілу. Друга – однорідність дисперсій між групами (homogeneity of variance). Тобто, крім нормального розподілу в кожній групі, значення залежної змінної повинні мати однаковий ступінь розкиду. Це важливо, оскільки при значній різниці у групових дисперсіях застосовувані формули можуть некоректно працювати. Якщо ці умови не виконуються, потрібно використовувати непараметричні аналоги ANOVA.

Перевірка однорідності дисперсій може здійснюватися графічними та формальними методами. Візуально оцінити однорідність можна за допомогою діаграм розмахів (якщо висота «ящиків» значно відрізняється, дисперсія неоднорідна). Формально однорідність перевіряється тестами, зокрема тестом Левене (Levene's test). Якщо р-рівень цього тесту перевищує 0,05, дисперсії можна вважати однорідними. Якщо ж значення менше 0,05, нульова гіпотеза про рівність дисперсій відкидається. Також вибірки повинні бути незалежними, а спостереження в межах кожної вибірки – несуміжними.

**Аналіз залишків (residuals).** Залишки характеризують внутрішньогрупову дисперсію, яка не може бути пояснена впливом експериментального фактора і є наслідком випадкової флуктуації або дії іншого фактора.

**Інтерпретація підсумкової таблиці дисперсійного аналізу.** На початковому етапі аналізу, незалежно від використовуваного програмного пакету, отримують таблицю, в якій ключовим є р-рівень. Якщо його значення < 0,05, вибірки статистично значимо відрізняються.

**Основні показники:**

- SS (Sum of Squares) – сума квадратів відхилень у групах,
- MS (Mean of Squares) – оцінка дисперсії між групами,
- F – критерій Фішера (оцінка різниці між внутрішньогруповою та міжгруповою дисперсіями),

- df (degrees of freedom) – число ступенів свободи,
- p – p-рівень.

**Апостеріорні порівняння (post-hoc tests).** Дисперсійний аналіз лише встановлює наявність або відсутність відмінностей між вибірками, але не показує, які саме групи відрізняються. Для цього використовують попарні порівняння середніх значень. Один із найпопулярніших методів – критерій Тьюкі (Tukey's HSD test). Якщо значення p-рівня менше 0,05, відмінності між групами є достовірними.

Для коригування похибки першого роду часто застосовують поправку Бонферроні (Bonferroni correction), за якої критичний рівень значущості ділиться на кількість порівнянь. Наприклад, якщо здійснюється три порівняння, то новий критичний рівень дорівнює  $0,05 / 3 = 0,017$ .

**Умови застосування непараметричного однофакторного дисперсійного аналізу.** Критерій Краскела–Уолліса (Kruskal–Wallis H-test) – непараметричний аналог однофакторного дисперсійного аналізу, що використовується для аналізу вибірок, які не відповідають закону нормального розподілу або мають неоднорідні дисперсії. У таких випадках дані перетворюють у ранги для подальшого аналізу.

Подібно до класичного ANOVA, тест Краскела–Уолліса дозволяє визначити, чи є статистично значущі відмінності між вибірками ( $p < 0,05$ ) або ж таких відмінностей немає ( $p > 0,05$ ). Для визначення конкретних груп, що різняться, проводяться додаткові апостеріорні тести, наприклад критерій Данна (Dunn's post hoc test).

Критерій Фрідмана – непараметричний аналог однофакторного дисперсійного аналізу для залежних вибірок, що містять повторні вимірювання, наприклад, якщо дані отримані з тих самих об'єктів у різний час.

**Виконання аналізу.** Для прикладу, в таблиці вихідних даних наведено висоту трьох видів чагарничків (багна, хамедафни та лохини), виміряну в 20-кратній повторності. Необхідно визначити, чи відрізняється висота цих рослин однієї життєвої форми, що ростуть в одному типі лісу.

**Розвідувальний аналіз даних.** Завантажуємо таблицю з Excel у PAST 4.03 та перевіряємо відповідність нормальному розподілу (Univariate/Normality tests). Оскільки частина даних не задовольняє умову нормальності виконуємо логарифмічну трансформацію (Transform – Log). Повторна перевірка нормальності (тест Шапіро–Вілка,  $p > 0,05$ ) підтвердила відповідність даних нормальному розподілу.

The screenshot shows the PAST 4.03 interface with two data tables and two windows for normality tests. The top table shows raw data for three species across six samples. The bottom table shows the data after a log transformation. The test windows display Shapiro-Wilk W and p(normal) values for each species.

	Ledum palustre	Chamaedaphne calyculata	Vaccinium uliginosum
1	6,9900147898	6,8855080871	6,1690443261
2	6,3678274305	6,0456593359	6,2094653672
3	6,4066242644	5,9189037258	5,6084589037
4	6,4833134834	6,0037945204	6,7428657418
5	6,6700482328	6,7065857199	6,3287202277
6	6,3678274305	5,8323965824	6,8502092551

	Ledum palustre	Chamaedaphne calyculata	Vaccinium uliginosum
1	0,8444780946	0,8379359927	0,7902178907
2	0,8039912854	0,7814436716	0,7930542091
3	0,8066292542	0,7722412759	0,7488434954
4	0,8117970212	0,7784258201	0,8288445124
5	0,8241289744	0,8265014798	0,8013158973

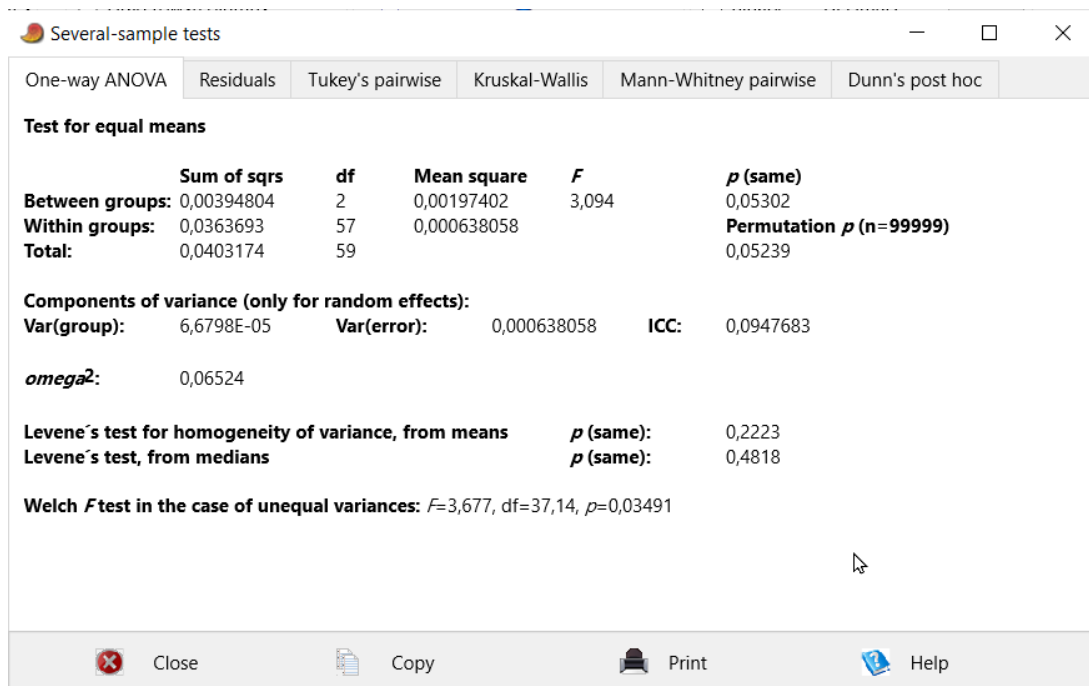
	Ledum palustre	Chamaedaphne calyculata	Vaccinium uliginosum
N	20	20	20
Shapiro-Wilk W	0,9366	0,9018	0,948
p(normal)	0,207	0,04459	0,3378

	Ledum palustre	Chamaedaphne calyculata	Vaccinium uliginosum
N	20	20	20
Shapiro-Wilk W	0,9434	0,9058	0,9512
p(normal)	0,2782	0,05311	0,3851

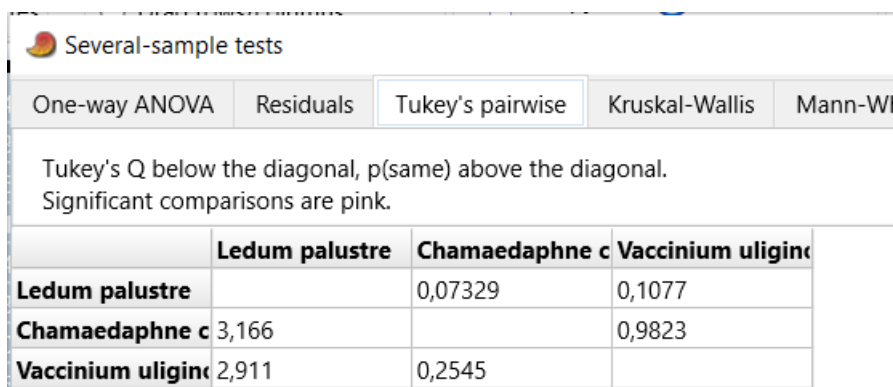
Далі потрібно переконатися, що виконується умова однорідності дисперсій. Для цього слід скористатися вкладкою Univariate/ANOVA etc. (several samples)/Several sample tests (ANOVA,

Kruskal–Wallis)/ANOVA (власне, це і є панель дисперсійного аналізу) та звернути увагу на значення тесту Левене (Levene’s test). У нашому прикладі р-рівень цього тесту перевищує 0,05 ( $p = 0,22$ ). Отже, і друга умова виконується. Тому ми можемо провести параметричний дисперсійний аналіз (one-way ANOVA).

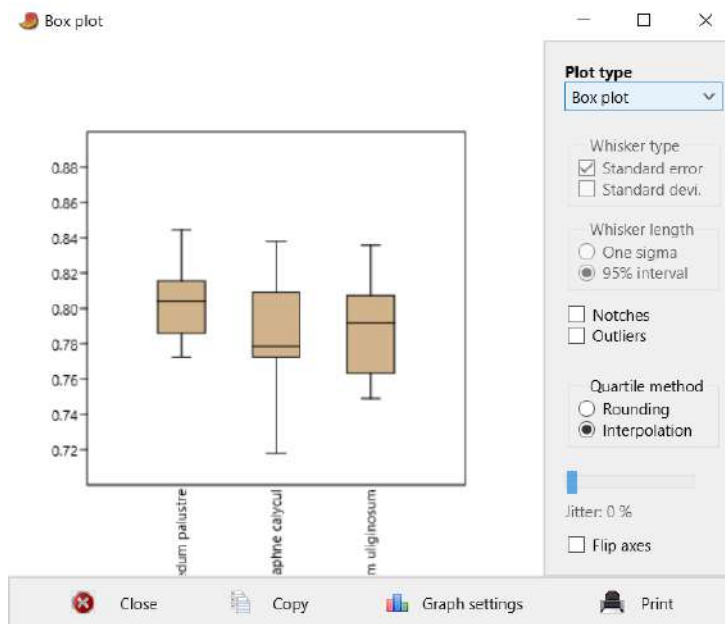


**Дисперсійний аналіз.** Використовуючи ту ж вкладку Univariate/ANOVA etc. (several samples)/Several sample tests (ANOVA, Kruskal–Wallis)/ANOVA, аналізуємо таблицю результатів Test for equal means. Як видно з рис. 2,  $F = 3,09$ ,  $p = 0,053$ . Отже, достовірних відмінностей між висотою чагарничків немає. Перестановочний тест (Permutation test) підтверджує ці результати ( $p = 0,052$ ).

**Апостеріорні порівняння.** Виконуються за допомогою вкладки Univariate/ANOVA etc. (several samples)/Several sample tests (ANOVA, Kruskal–Wallis)/Tukey’s pairwise. Тест Тьюкі також не виявив значущих відмінностей між значеннями конкретних змінних при їх попарних порівняннях:  $p > 0,05$

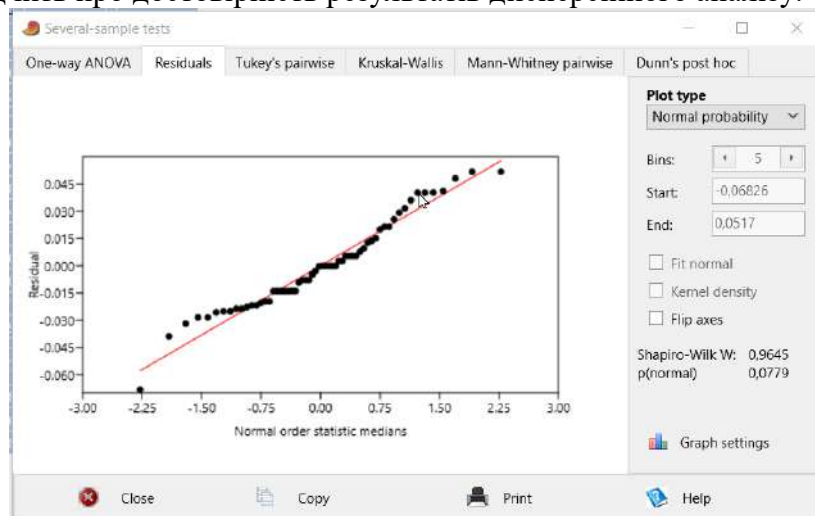


**Візуалізація результатів аналізу.** Оскільки дисперсійний аналіз лише виявляє наявність або відсутність відмінностей між змінними, але не відображає їхню величину, розподіл змінних можна зобразити графічно. Найчастіше використовуються діаграми розмаху («ящики з вусами»). Їх можна побудувати за допомогою вкладки Plot/Barchart/Box plot.



Як видно з рис., на перший погляд, відмінності є, принаймні за розмірами «ящиків», але вони статистично незначущі. Крім того, якщо подивитися на «вуса», то вони розташовані приблизно на одному рівні, що також свідчить про відсутність значущих відмінностей.

**Аналіз залишків.** Для оцінки достовірності проведеного аналізу досліджуються залишки моделі, які повинні відповідати закону нормального розподілу. Це перевіряється за допомогою вкладки Residuals. Як видно з діаграми розсіювання та значення тесту Шапіро–Вілка ( $p > 0,05$ ), дана умова виконується, що свідчить про достовірність результатів дисперсійного аналізу.



**Однофакторний дисперсійний аналіз (Kruskal–Wallis test) для непараметричних даних.** У таблиці вихідних даних, які представлені у стовпцях, вказано кількість видів гіпотетичних тварин, зареєстрованих у трьох біотопах за результатами 10 обліків у кожному з них. Необхідно визначити, чи відрізняється видове багатство цих трьох місцевіснвань.

**Розвідувальний аналіз даних.** Завантажуємо таблицю з даними з файлу MS Excel у PAST 4.04 та виконуємо перевірку на нормальність розподілу за допомогою вкладки Univariate/Normality tests. Перевірка показала, що деякі дані наближені до нормального розподілу, але не відповідають йому. Після логарифмування ситуація не змінилася, що підтверджується відповідними тестами, зокрема тестом Шапіро–Вілка ( $p < 0,05$ ) (рис.). Отже, необхідно виконати непараметричний дисперсійний аналіз. Оскільки дані не відповідають закону нормального розподілу, перевірка на однорідність дисперсій не потрібна.

	біотоп 1	біотоп 2	біотоп 3
1	12	10	24
2	11	10	24
3	13	11	22
4	12	11	24
5	12	10	22
6	14	10	24
7	13	11	22
8	12	10	24
9	11	11	22
10	12	10	24

	біотоп 1	біотоп 2	біотоп 3
N	10	10	10
Shapiro-Wilk W	0,885	0,6405	0,6405
p(normal)	0,1488	0,0001687	0,0001687
Anderson-Darling A	0,6281	1,737	1,737
p(normal)	0,07186	7,326E-05	7,326E-05
p(Monte Carlo)	0,0767	0,0001	0,0003
Lilliefors L	0,2861	0,3807	0,3807
p(normal)	0,02048	0,0001	0,0001
p(Monte Carlo)	0,0212	0,0002	0,0004
Jarque-Bera JB	0,4717	1,678	1,678
p(normal)	0,7899	0,4321	0,4321
p(Monte Carlo)	0,7051	0,0939	0,1006

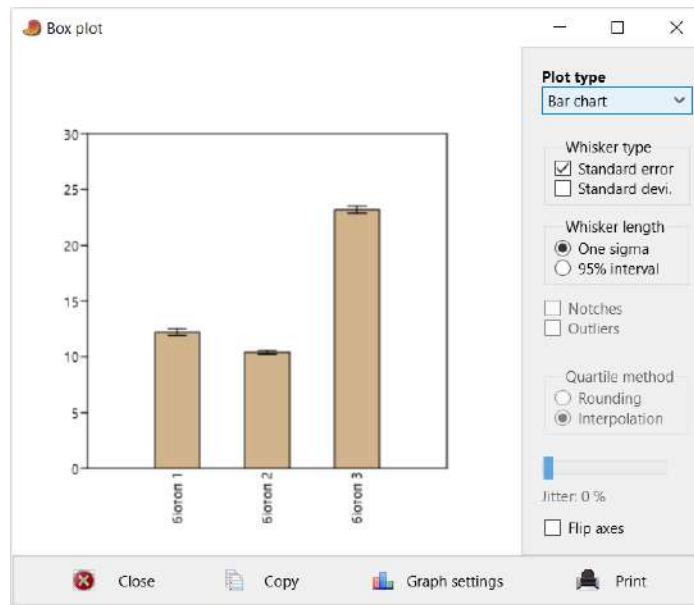
**Дисперсійний аналіз.** Для проведення дисперсійного аналізу використовується вкладка Univariate/ANOVA etc. (several samples)/Several sample tests (ANOVA, Kruskal–Wallis)/Kruskal–Wallis. Як видно з рис., значення тесту Краскела–Уолліса ( $H = 24,82$ ,  $p < 0,05$ ) свідчать про наявність значущих відмінностей у видовому багатстві трьох досліджуваних біотопів.

Several-sample tests	
One-way ANOVA	Residuals
Tukey's pairwise	Kruskal-Wallis
<b>Kruskal-Wallis test for equal medians</b>	
$H(ch^2)$ :	24,82
$H_c$ (tie corrected):	25,59
$p$ (same):	2,775E-06
There is a significant difference between sample medians	

**Апостеріорні порівняння.** Виконуються за допомогою вкладки Univariate/ANOVA etc. (several samples)/Several sample tests (ANOVA, Kruskal–Wallis)/Dunn's post hoc. Тест Данна виявив значущі відмінності між значеннями всіх змінних при їх попарних порівняннях:  $p < 0,05$ .

Several-sample tests			
One-way ANOVA	Residuals		
Tukey's pairwise	Kruskal-Wallis		
Mann-Whitney pairwise	Dunn's post hoc		
Raw p values, uncorrected significance			
	біотоп 1	біотоп 2	біотоп 3
біотоп 1		0,01765	0,007308
біотоп 2	0,01765		4,293E-07
біотоп 3	0,007308	4,293E-07	

**Візуалізація результатів аналізу.** Оскільки дисперсійний аналіз лише виявляє наявність або відсутність відмінностей між змінними, але не відображає їхньої величини, варіацію змінних можна представити на діаграмах розмаху, побудувавши їх у вкладці Plot/Barchart (рис.).



Як видно з рис., відмінності встановлено: у біотопі 3 видовий склад є достовірно вищим, а в біотопі 2 – найнижчим.

**Підготуйте звіт за результатами виконання роботи (з результатами аналізу та графіками в книзі Excell) та завантажте на гугл-диск для перевірки.**

#### Рекомендована література

1. Серікова О.М. Методи обробки статистичних даних : курс лекцій. – Х.: НУЦЗУ, 2019. – 198 с.
2. Янковська Л.В. Математичні методи в екології: навчальний посібник для вузів – Тернопіль: ТНПУ, 2017. – 114с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Тарасова В.В., Парфенцева Н.О., Ковалевська І.М. Екологічна статистика. Підручник. – К.: Центр навчальної літератури, 2013. – 295 с.
5. PAST: Paleontological Statistics Software Package for Education and Data Analysis. [https://palaeo-electronica.org/2001\\_1/past/past.pdf](https://palaeo-electronica.org/2001_1/past/past.pdf)

#### Лабораторна робота № 4

**Тема:** Статистичні методи редукції та класифікації даних.

**Мета:** навчитися використовувати метод аналізу головних компонент для зменшення розмірності даних.

#### Теоретичні питання

1. Сутність і призначення методів редукції даних.
2. Метод головних компонент.
3. Факторний аналіз.
4. Сутність кластерного аналізу.
5. Види кластерного аналізу.
6. Особливості застосування кластерного аналізу.

#### Хід роботи

**Сутність методу аналізу головних компонент (РСА).** Аналіз головних компонент (РСА) – це статистичний метод, який використовується для зменшення розмірності даних, виявляючи найбільш

значущі закономірності в багатовимірних даних. Він перетворює вихідні змінні в новий набір змінних, які називаються головними компонентами (PC). Ці головні компоненти є лінійними комбінаціями вихідних змінних і розташовані в порядку зменшення їхньої варіації.

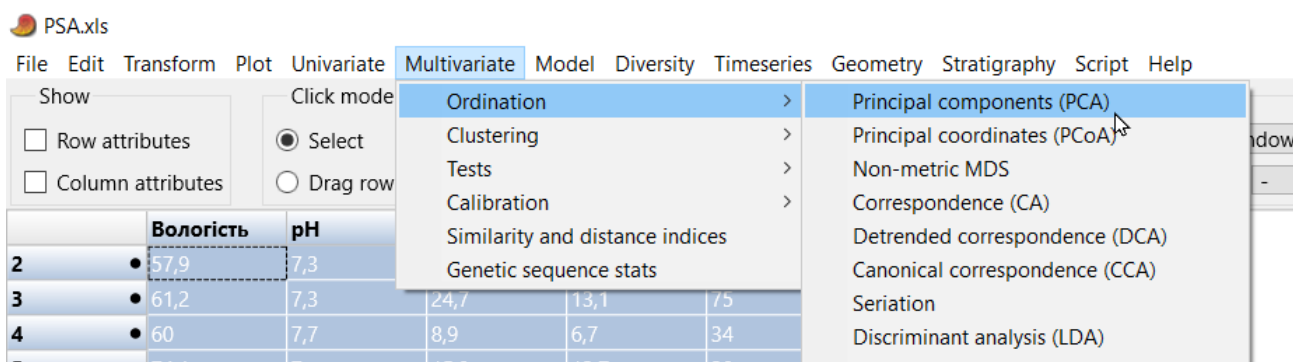
PCA широко використовується в екології для:

- **Виявлення основних факторів, що впливають на екосистеми:** PCA допомагає визначити, які екологічні фактори (наприклад, температура, вологість, забруднення) найбільше впливають на розподіл видів або стан екосистеми.
- **Класифікації екосистем:** PCA може використовуватися для класифікації екосистем на основі їхніх характеристик.
- **Аналізу біорізноманіття:** PCA допомагає виявити закономірності в розподілі видів і оцінити біорізноманіття.
- **Оцінки впливу забруднення:** PCA може використовуватися для оцінки впливу забруднення на екосистеми.
- **Моніторингу екологічних змін:** PCA допомагає відстежувати зміни в екосистемах з часом.

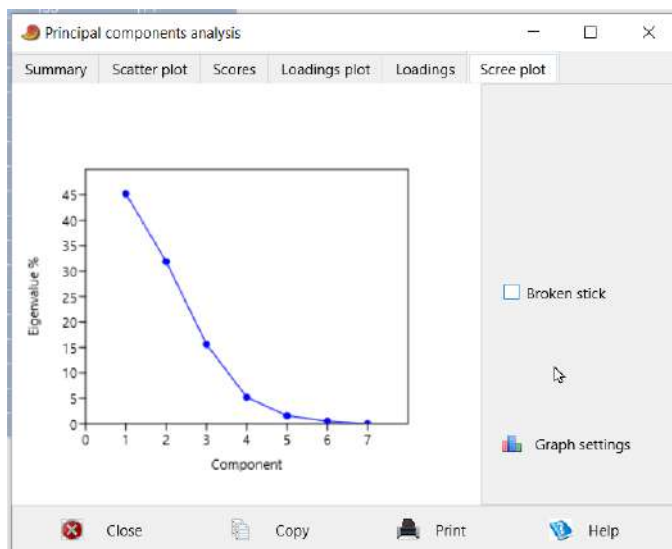
**Виконання аналізу.** Відкрийте PAST 4.03 і введіть навчальні дані в таблицю. Виділіть діапазон даних для аналізу.

№	Вологість	pH	Забрудн 1	Забрудн 2	Вид 1	Вид 2	Вид 3
1	51,5	7,2	20,2	7,7	37	19	61
2	57,9	7,3	12,2	3	73	31	26
3	61,2	7,3	24,7	13,1	75	33	71
4	60	7,7	8,9	6,7	34	30	32
5	74,1	7	15,8	13,7	28	18	13
6	79,3	7,6	21,5	13,9	70	57	36
7	59,4	6,5	13,6	2,4	10	52	54
8	68,5	6,9	24,3	4,1	75	53	42
9	57,8	6,6	20,6	5,5	25	72	57
10	51,2	6,5	5,9	2,1	25	59	39
11	67,8	7,4	12	12,2	28	30	48
12	61,7	7,6	11,7	9,3	60	65	53
13	58,9	7,8	12,4	4,6	52	25	58
14	63,8	6,9	17	10,7	11	62	78
15	67,8	7,2	9,9	14,1	20	58	52
16	79,8	7,7	13	13,8	72	79	34
17	61,1	6,6	8,2	12,7	69	23	25
18	68,2	7,6	19,7	2,8	27	47	48
19	57,4	7,1	18,5	4,8	34	21	58
20	77,7	7,4	15,5	8,5	30	18	36

Виберіть аналіз PCA: у меню "Multivariate" виберіть "Principal components" та виконайте аналіз.



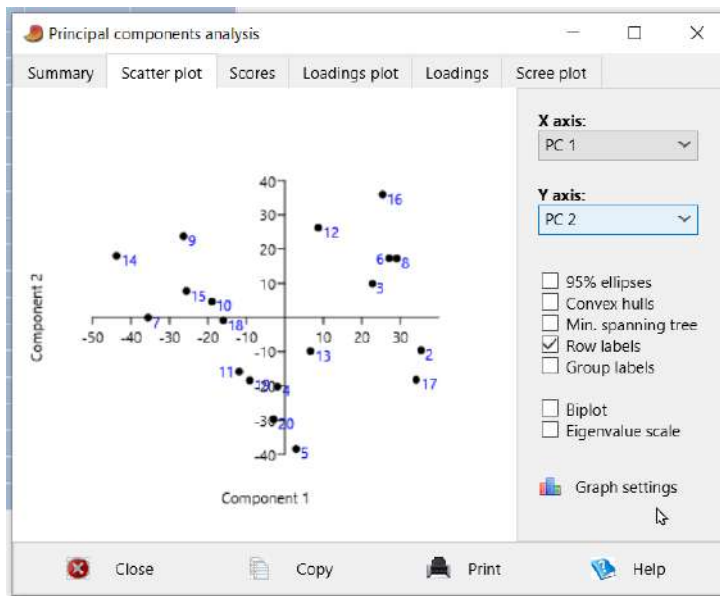
Перше, на що потрібно звернути увагу – це кількість головних компонент, які варто розглядати при подальшому аналізі результатів. Для цього перейдіть у вкладку Scree plot. Для подальшого аналізу мають значення головні компоненти, що лежать на «схилі» графіка. Тобто у нашому випадку можна обмежитись першими чотирма головними компонентами.



У вкладці Summary можна побачити оцінки внесків головних компонент у пояснення мінливості в межах аналізованого масиву даних. Як бачимо, перша головна компонента пояснює понад 45% дисперсії, а чотири перших головні компоненти пояснюють майже 98% дисперсії. Це означає, що всі інші головні компоненти пояснюють близько 2% загальної дисперсії, що дозволяє їх виключити з подальшого аналізу.

PC	Eigenvalue	% variance
1	593.123	45.221
2	418.046	31.873
3	204.693	15.506
4	68.1088	5.1928
5	20.7922	1.5853
6	6.741	0.51395
7	0.0927837	0.0070741

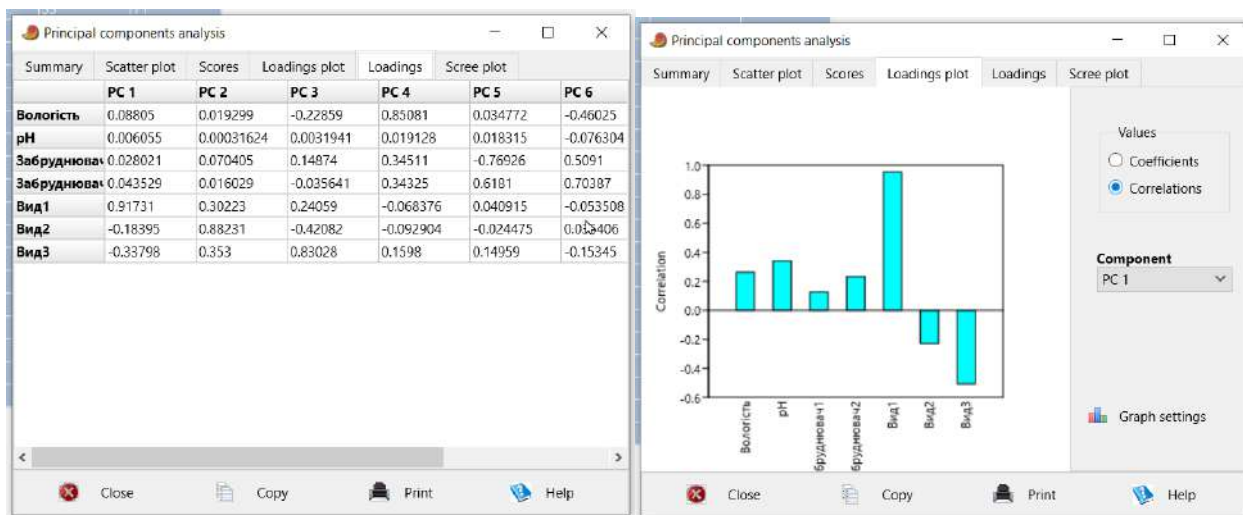
У вкладці Scatter plot представлена діаграма розсіювання досліджуваних випадків у просторі головних компонент. Оскільки простір графіка двохвимірний, ви можете по черзі переглянути положення досліджуваних випадків у полі різних пар головних компонент. На графіку розсіювання у полі перших двох ГК добре помітно, що аналізовані випадки утворюють декілька достатньо чітких груп, які можна інтерпретувати на основі ГК.



У вкладці Scores представлені значення головних компонент для кожного випадку. Оскільки ми уже знаємо, що найбільше значення у нашому випадку мають перші чотири ГК, ми можемо використати ці дані для подальшого аналізу, редукувавши таким чином кількість змінних. Ці дані також можна використати в якості предикторів для потреб екологічного моделювання.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
2	35.415	-9.4991	-2.0083	-12.707	-2.6999	-1.0655
3	22.754	9.8606	35.739	4.7496	0.80621	3.9485
4	-1.9477	-20.183	-7.0903	-7.0632	1.5323	-0.0090
5	2.9127	-38.421	-21.708	8.1927	-1.7653	4.7944
6	27.122	17.307	-9.2646	11.844	-1.6303	1.0411
7	-35.561	-0.0092368	-2.8706	-4.3378	-3.0134	-1.566
8	29.113	17.238	1.8356	1.2334	-9.03	-0.7410
9	-26.304	23.741	-3.8905	-4.6217	-5.9629	4.3736
10	-18.97	4.7001	-13.922	-18.148	0.63796	-0.1691
11	-11.848	-15.891	3.2319	5.4921	4.9608	-0.2607
12	8.7073	26.241	1.8072	-5.4336	4.3912	-0.8677
13	6.6067	-9.7845	21.779	-4.1216	2.2536	-4.3416
14	-43.748	18.046	12.294	6.2734	3.0479	0.5791
15	-25.666	7.6902	-7.5354	4.0006	7.3327	0.8563
16	25.388	36.026	-21.077	6.8038	4.11	-2.5935

В таблиці у вкладці Loadings представлені оцінки кореляцій головних компонент з аналізованими параметрами. У вкладці Loadings plot ці дані представлено у графічному форматі. На основі цих оцінок ми можемо інтерпретувати головні компоненти. Так ми бачимо, що перша ГК має найвищий коефіцієнт кореляції (0,92) з параметром Вид 1. Отже перша ГЛ, яка пояснює 45% загальної дисперсії, найбільше навантажується цим параметром. Друга ГК має максимальну кореляцію з параметром Вид 2, третя ГК – з параметром Вид 3 а четверта ГК з параметром вологості.



Отже, результати проведеного аналізу свідчать, що серед усього аналізованого масиву даних найбільше значення мають перші чотири ГК, які добре корелюють з параметрами «Вид 1»-«Вид 2» та «Вологість». А отже для подальшого аналізу ми можемо залишити лише ці чотири параметри, або замінити їх значення значеннями головних компонент.

**Кластерний аналіз** — це метод машинного навчання без учителя, який використовується для групування схожих об'єктів у кластери. Основна мета — виявити приховані закономірності та структури в даних, розділяючи їх на групи на основі їх схожості.

Алгоритми кластерного аналізу в PAST 4.03

**Ієрархічний кластерний аналіз (Hierarchical Clustering).** Цей метод будує ієрархію кластерів, починаючи з кожного об'єкта як окремого кластера. На кожному кроці два найближчі кластери об'єднуються. Результати відображаються у вигляді дендрограми, яка показує ієрархічну структуру кластерів. Висота з'єднання кластерів відображає відстань між ними. Ієрархічний кластерний аналіз може показати, які екосистеми найбільш схожі за своїми екологічними характеристиками. Наприклад, тропічний ліс і кораловий риф можуть бути об'єднані в один кластер, оскільки вони мають високе видове різноманіття, щільність популяції та температуру.

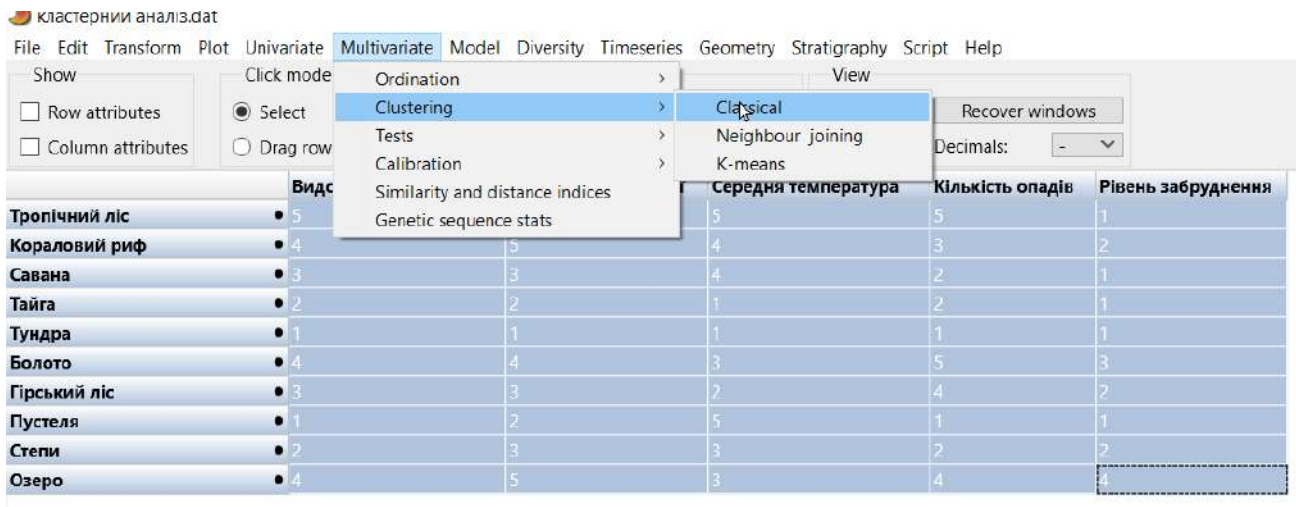
**К-середніх (K-means).** Цей метод розділяє дані на задану кількість (k) кластерів. Алгоритм ітеративно присвоює кожному об'єкту найближчий центр кластера і перераховує центри кластерів до збіжності. Результати показують, до якого кластера належить кожен об'єкт. Центри кластерів відображають середні значення характеристик для кожного кластера. К-середніх може розділити екосистеми на задану кількість груп. Наприклад, при k=3 можна отримати кластери з високим різноманіттям (тропічний ліс, кораловий риф, болото, озеро), середнім різноманіттям (савана, гірський ліс, степи) та низьким різноманіттям (тайга, тундра, пустеля).

**Проведення аналізу.** Для проведення аналізу використаємо навчальні дані з таблиці.

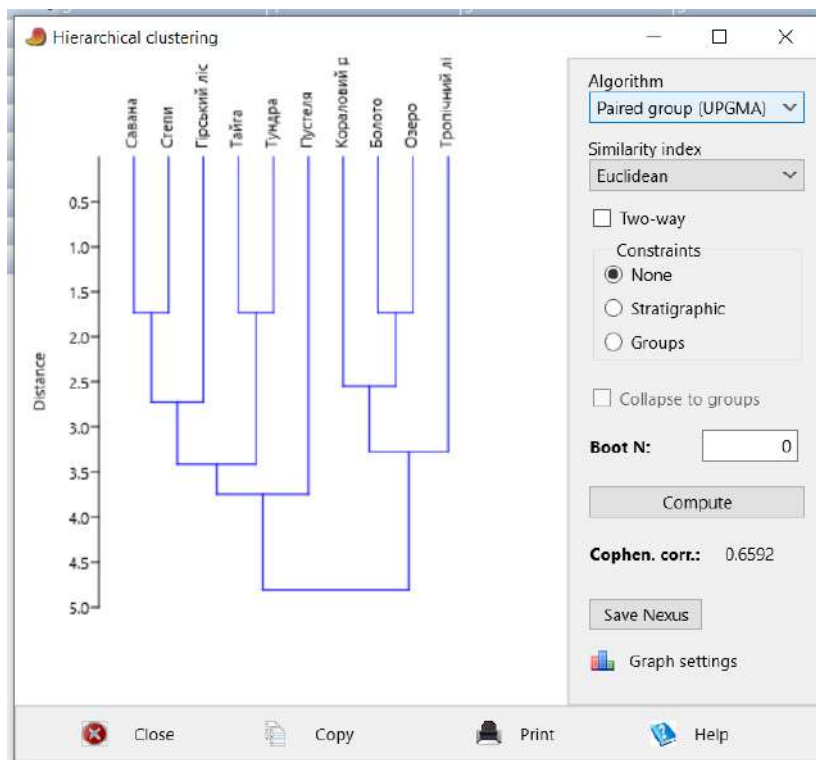
Екосистема	Видове різноманіття	Щільність популяції	Середня температура	Кількість опадів	Рівень забруднення
Тропічний ліс	5	4	5	5	1
Кораловий риф	4	5	4	3	2
Савана	3	3	4	2	1
Тайга	2	2	1	2	1
Тундра	1	1	1	1	1
Болото	4	4	3	5	3
Гірський ліс	3	3	2	4	2

Пустеля	1	2	5	1	1
Степи	2	3	3	2	2
Озеро	4	5	3	4	4

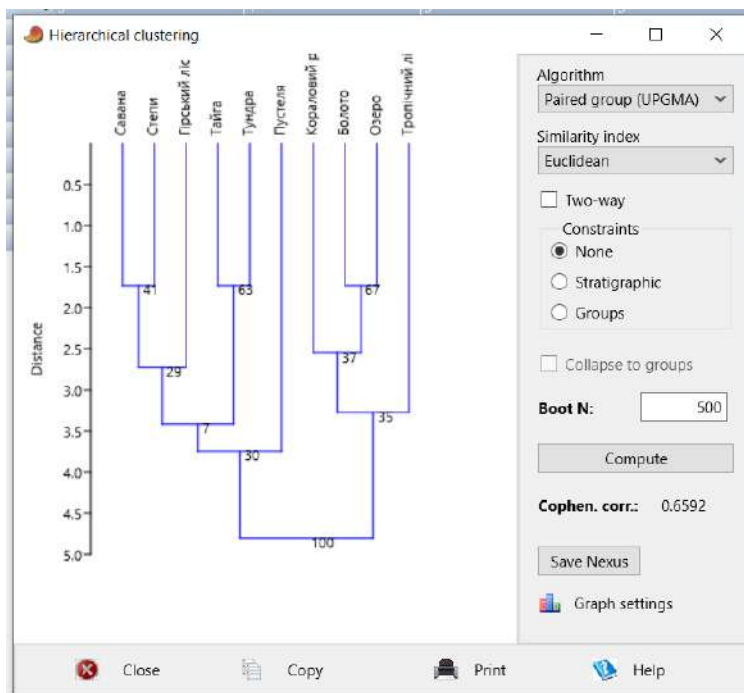
Введіть дані в PAST у вигляді таблиці. Виділіть весь масив даних для аналізу та перейдіть у вкладку Multivariate – Clustering – Classical, щоб виконати **деревоподібну кластеризацію**.



Отримана дендрограма демонструє наявність кількох груп у наших даних. У програмному пакеті доступні різні алгоритми кластеризації та індекси подібності. Ви можете поекспериментувати з ними і відібрати оптимальний результат, відповідний завданням аналізу. Однак у такому вигляді ми не можемо оцінити достовірність результату кластеризації.



Для оцінки вірогідності утворення кластерів використовується бутстреп-аналіз. Його суть полягає в тому, що програма генерує певну кількість дерев класифікації, а потім оцінює, з якою частотою повторюється той чи інший кластер у цьому масиві результатів. Наприклад, якщо ми задамо 500 ітерацій для бутстреп-аналізу, ми отримаємо такий результат:

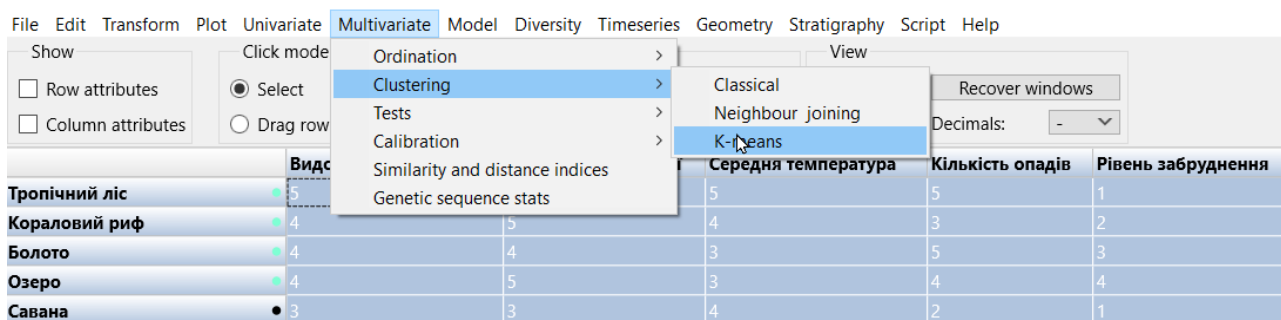


Результати бутстреп-оцінки свідчать про те, що лише два кластери утворюються з високою вірогідністю (100%). Наступні кластери утворюються з вірогідністю менше 50%, а отже можна стверджувати, що вони є випадковими. На цьому етапі можна інтерпертувати кластери. Для цього буває корисним порівняти вихідні дані для отриманих груп.

	Color	Symbol	Name	Видове різноманіття	Щільність популяції	Середня температура	Кількість опадів	Рівень забруднення
Тропічний ліс	Aquamarine	Dot	Тропічний ліс	5	4	5	5	1
Кораловий риф	Aquamarine	Dot	Кораловий риф	4	5	4	3	2
Болото	Aquamarine	Dot	Болото	4	4	3	5	3
Озеро	Aquamarine	Dot	Озеро	4	5	3	4	4
Савана		Dot	Савана	3	3	4	2	1
Тайга		Dot	Тайга	2	2	1	2	1
Тундра		Dot	Тундра	1	1	1	1	1
Гірський ліс		Dot	Гірський ліс	3	3	2	4	2
Пустеля		Dot	Пустеля	1	2	5	1	1
Степи		Dot	Степи	2	3	3	2	

Як видно з перегрупованих даних відповідно до результатів кластеризації, екосистеми другого кластеру характеризуються вищими значеннями показників біорізноманіття, а до першого кластеру увійшли екосистеми з низькими оцінками біорізноманіття (видове різноманіття та щільність популяції). Якщо розглянути абіотичні параметри, ми можемо зробити висновки про причини таких відмінностей. Так екосистеми другого кластера існують в умовах вищих температур та вологості, що і може бути причиною вищої видової різноманітності та чисельності живих організмів.

**Метод кластеризації К-середніх.** Виділіть масив даних для аналізу, перейдіть у вкладку Multivariate – Clustering – K-means та запустіть аналіз.



У наступному вікні потрібно задати кількість кластерів. Використаємо поділ на два кластери, щоб була можливість порівняння результатів двох методів кластеризації.



Item	Cluster
Тропічний ліс	1
Кораловий риф	1
Болото	1
Озеро	1
Савана	2
Тайга	2
Тундра	2
Гірський ліс	1
Пустеля	2
Степи	2

Як бачимо, алгоритм розділив аналізовані дані на два кластери, структура яких майже ідентична результатам деревоподібної кластеризації. Кожен з використаних методів має свої переваги і недоліки.

Деревоподібна (ієрархічна) кластеризація надає візуалізацію ієрархії кластерів, що дозволяє досліджувати структуру даних на різних рівнях, не вимагає попереднього визначення кількості кластерів, добре підходить для даних з ієрархічною структурою. Серед недоліків цього методу варто відзначити те, що він обчислювально витратний для великих наборів даних та чутливий до вибору метрики відстані.

Метод К-середніх (K-means) обчислювально ефективний для великих наборів даних, простий у реалізації та добре підходить для даних з компактними та відокремленими кластерами. Недоліки цього методу: вимагає попереднього визначення кількості кластерів (k), чутливий до початкового вибору центрів кластерів, не гарантує знаходження глобального оптимуму, погано обробляє не компактні кластери. Ці особливості потрібно враховувати для прийняття рішення про використання того чи іншого алгоритму кластеризації.

**Підготуйте звіт за результатами виконання роботи (з результатами аналізу та графіками в книзі Excell) та завантажте на гугл-диск для перевірки.**

#### Рекомендована література

1. Серікова О.М. Методи обробки статистичних даних : курс лекцій. – Х.: НУЦЗУ, 2019. – 198 с.
2. Янковська Л.В. Математичні методи в екології: навчальний посібник для вузів – Тернопіль: ТНПУ, 2017. – 114с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Тарасова В.В., Парфенцева Н.О., Ковалевська І.М. Екологічна статистика. Підручник. – К.: Центр навчальної літератури, 2013. – 295 с.
5. PAST: Paleontological Statistics Software Package for Education and Data Analysis. [https://palaeo-electronica.org/2001\\_1/past/past.pdf](https://palaeo-electronica.org/2001_1/past/past.pdf)

#### Лабораторна робота № 5

**Тема: Методи аналізу  $\alpha$ - та  $\beta$ -біорізноманіття.**

**Мета:** навчитись розраховувати індекси альфа-різноманіття та бета-різноманіття, здійснювати оцінку статистичної значущості їх відмінностей, будувати профілі різноманітності, оцінювати зв'язок видового багатства з об'ємом вибірки, будувати матрицю подібності.

### Теоретичні питання

1. Поняття про біорізноманіття.
2. Рівні біорізноманіття.
3. Методи аналізу а-різноманіття.
4. Методи аналізу б-різноманіття.

### Хід роботи

Для аналізу біорізноманіття використаємо навчальні дані щодо чисельності наземних молюсків у чотирьох вибірках з міста Житомира.

Species	1	2	3	4
<i>A.circumscriptus</i>	0	0	1	1
<i>A.distinctus</i>	1	1	0	0
<i>A.fasciatus</i>	2	3	0	0
<i>B.fruticum</i>	7	2	1	1
<i>C.vindobonensis</i>	9	8	7	4
<i>H.albescens</i>	22	9	1	8
<i>L.maximus</i>	0	1	3	0
<i>O.elegans</i>	2	0	0	1
<i>S.putris</i>	0	0	0	1

**1. Розрахунок індексів альфа-різноманіття.** Виділіть всі колонки з даними (за допомогою клавіші Shift). У команді меню Diversity виберіть пункт Diversity Indices. В отриманому вікні представлені найбільш уживані індекси альфа-різноманіття. Результати розрахунків можна скопіювати в буфер обміну (за допомогою кнопки Copy data) і вставити в документ Microsoft Office.

Numbers	Plot	1	2	3	4
Taxa_S		6	6	5	6
Individuals		43	24	13	16
Dominance_D		0.3369	0.2778	0.3609	0.3281
Simpson_1-D		0.6631	0.7222	0.6391	0.6719
Shannon_H		1.339	1.466	1.264	1.386
Evenness_e^H/S		0.6356	0.7219	0.7076	0.6667
Brillouin		1.171	1.204	0.9412	1.056
Menhinick		0.915	1.225	1.387	1.5
Margalef		1.329	1.573	1.559	1.803
Equitability_J		0.7471	0.8181	0.7851	0.7737
Fisher_alpha		1.896	2.568	2.975	3.487
Berger-Parker		0.5116	0.375	0.5385	0.5
Chao-1		6	6.5	8	12

**Оцінка статистичної значущості відмінностей індексів різноманітності.** Виділіть дві колонки (напр., Найбільш контрастні за різноманітністю). Виберіть пункт Diversity – Diversity permutation test. В отриманому вікні величина p (eq) вказує ймовірність того, що відмінності в обчислених індексах різноманітності випадкові.

	1	2	Perm p(eq)
<b>Taxa S</b>	6	6	1
<b>Individuals</b>	43	24	0
<b>Dominance</b>	0.3369	0.2778	0.4758
<b>Shannon H</b>	1.339	1.466	0.6001
<b>Evenness e<sup>H</sup>/S</b>	0.6356	0.7219	0.518
<b>Simpson indx</b>	0.6631	0.7222	0.4758
<b>Menhinick</b>	0.915	1.225	0.3569
<b>Margalef</b>	1.329	1.573	0.7274
<b>Equitability J</b>	0.7471	0.8181	0.4609
<b>Fisher alpha</b>	1.896	2.568	0.4709
<b>Berger-Parker</b>	0.5116	0.375	0.3025

Unbiased

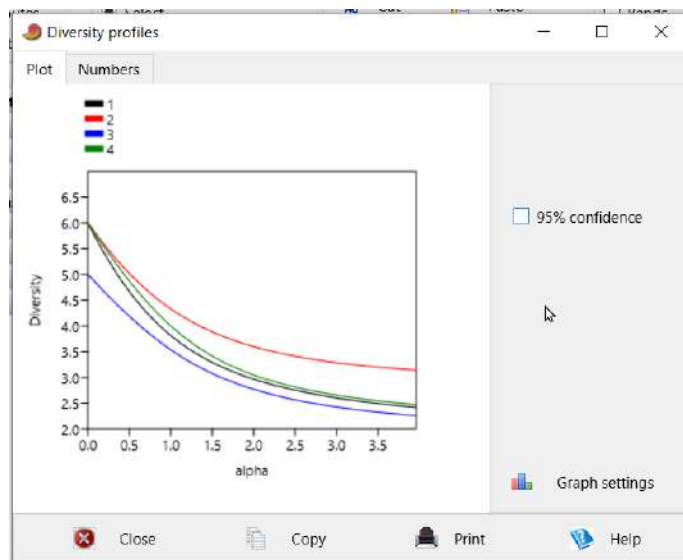
Close Copy Print Help

Значимість відмінностей індексів Шеннона та Симпсона за допомогою критерію Стюдента можна оцінити в пункті "Diversity t test".

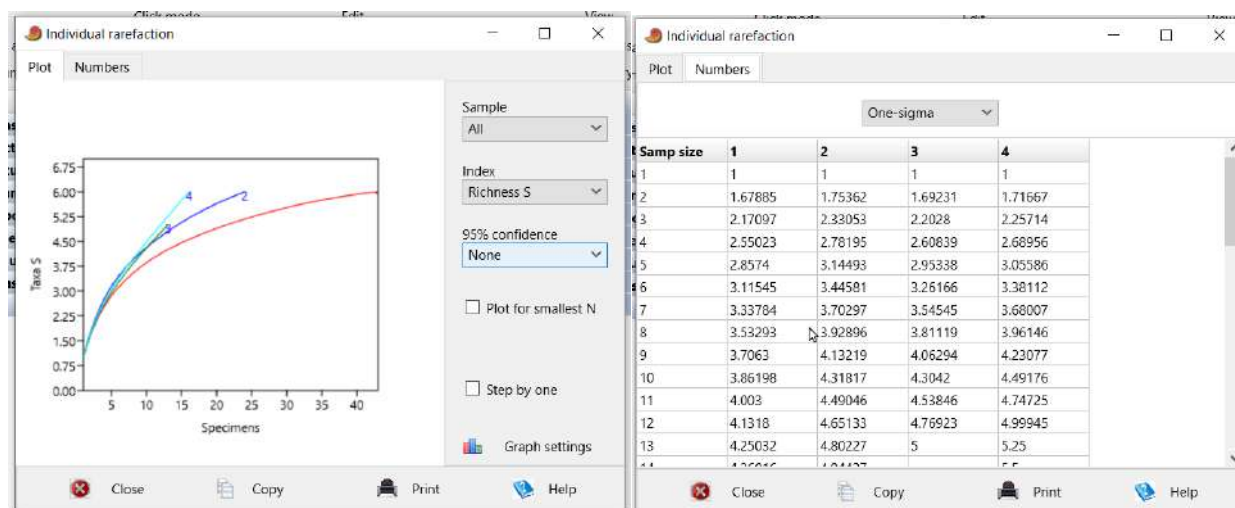
Shannon index			
1		2	
<i>H'</i>	1.3386	<i>H'</i>	1.4659
Variance:	0.017421	Variance:	0.025635
<i>t</i>	-0.61331		
df:	53.828		
<i>p</i> (same):	0.54225		
Simpson index			
<i>D</i>	0.33694	<i>D</i>	0.27778
Variance:	0.0033435	Variance:	0.003164
<i>t</i>	0.73338		
df:	62.543		
<i>p</i> (same):	0.46607		

Close Copy Print Help

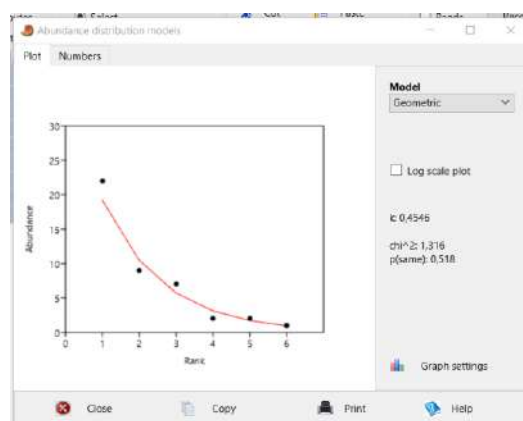
**Профілі різноманітності** легко побудувати за допомогою пункту Diversity profiles. Чим вище починається профіль, тим вище видове багатство, і чим ближче профіль до горизонтальної лінії, тим вище вирівняність угруповання. Якщо профілі для різних угруповань перетинаються, їх неможливо однозначно упорядкувати за видовою різноманітністю.



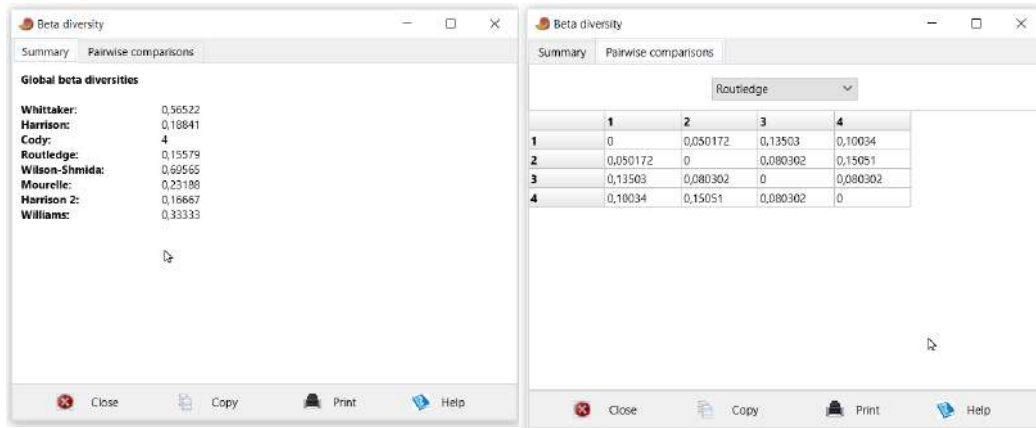
Оцінити зв'язок видового багатства з об'ємом вибірки можна за допомогою методу розрідження (Diversity - Individual rarefaction).



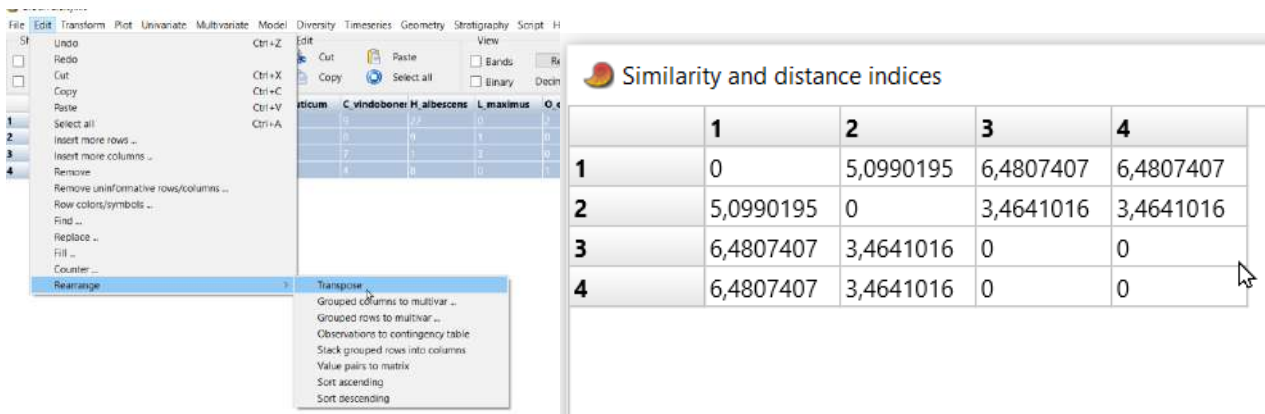
Моделі рангового розподілу видового різноманіття. Виділіть одну колонку. У команді меню Models виберіть пункт Abundance. Крім графіка, у вікні представлені параметри моделі. Значимість відмінностей оцінюється за критерієм хі-квадрат. Програма дозволяє тестувати чотири моделі розподілу: геометричну (geometric), логарифмічну (log series), логнормальну (log-normal), модель розламаного стержня (broken stick).



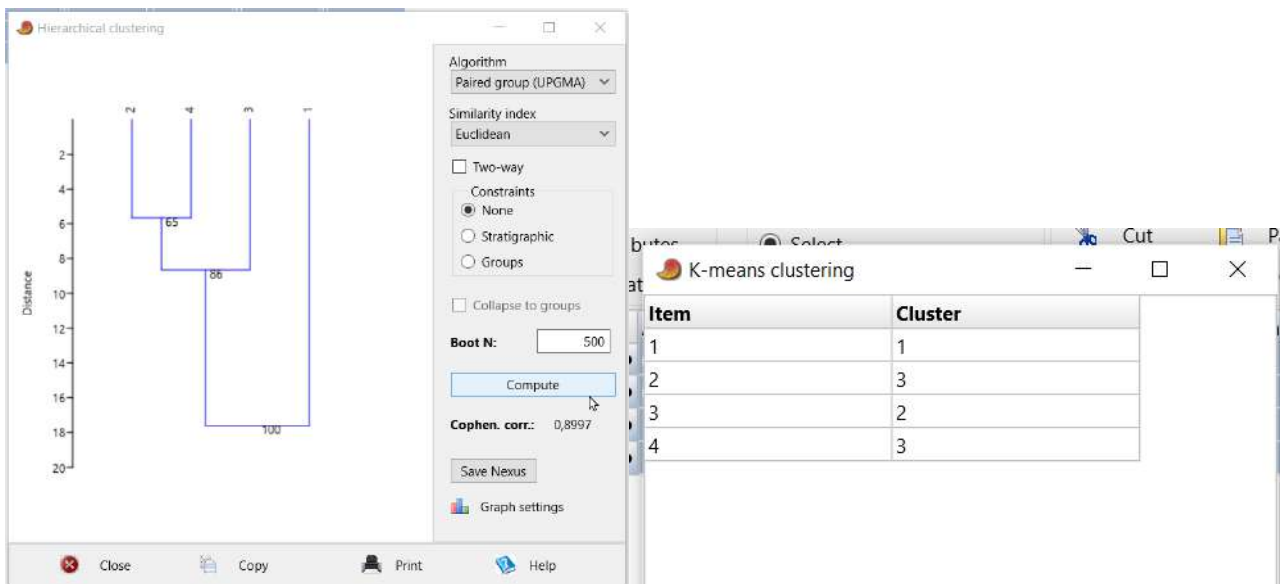
2. Бета-різноманітність можна виміряти в пункті Beta diversity меню Diversity.



**Матрицю подібності** можна побудувати за допомогою команди Similarity and distance indices в меню Multivariate. Тільки масив даних для цього потрібно транспонувати.

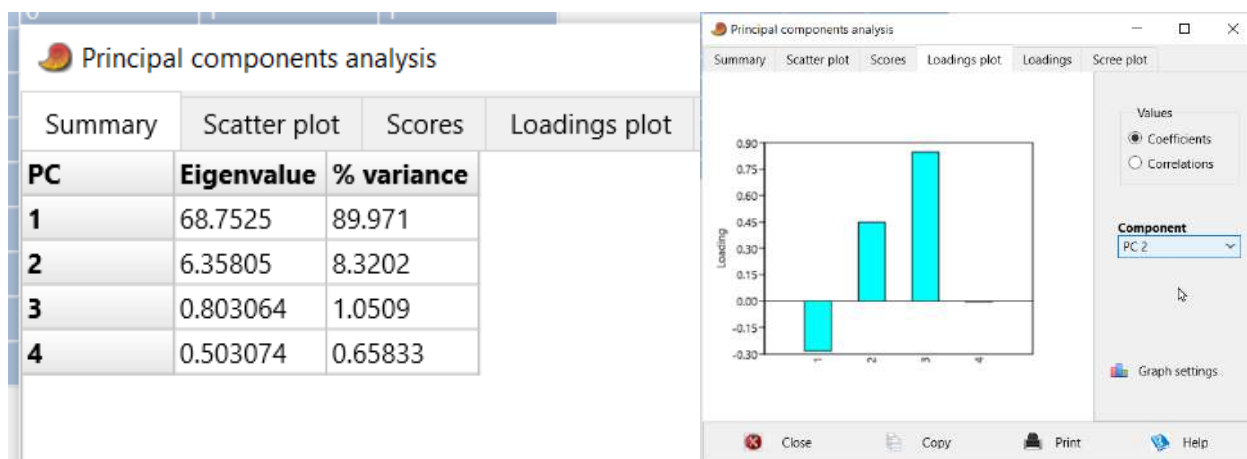


Класифікацію угруповань можна проводити за допомогою команд Cluster analysis (ієрархічний кластерний аналіз) або K-means (класифікація методом К-середніх) в меню Multivar (Багатомірний аналіз). У першому випадку програма виводить тільки Дендрограму, пропонується лише три алгоритми об'єднання.



**Ординацію угруповань** також можна виконати в меню Multivar. Реалізовано різні методи: аналіз відповідності (Correspondence), аналіз відповідності з віддаленим трендом (Detrended

correspondence), аналіз головних компонент (Principal components), аналіз головних координат (Principal coordinates), багатовимірне неметричне шкалювання (Non-metric MDS). У кожному методі можна вивести діаграму розсіювання (scatter), відомості про дисперсію (eigenvalue), що пояснюється кожної віссю та ін.



**Підготуйте звіт за результатами виконання роботи (з результатами аналізу та графіками в книзі Excell) та завантажте на гугл-диск для перевірки.**

#### Рекомендована література

1. Серікова О.М. Методи обробки статистичних даних : курс лекцій. – Х.: НУЦЗУ, 2019. – 198 с.
2. Янковська Л.В. Математичні методи в екології: навчальний посібник для вузів – Тернопіль: ТНПУ, 2017. – 114с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя : КПУ, 2011. – 268 с.
4. Тарасова В.В., Парфенцева Н.О., Ковалевська І.М. Екологічна статистика. Підручник. – К.: Центр навчальної літератури, 2013. – 295 с.
5. PAST: Paleontological Statistics Software Package for Education and Data Analysis. [https://palaeo-electronica.org/2001\\_1/past/past.pdf](https://palaeo-electronica.org/2001_1/past/past.pdf)

#### Лабораторна робота №6

**Тема: Основи роботи з просторовими даними в Q-GIS. Створення макету карти та підготовка до друку.**

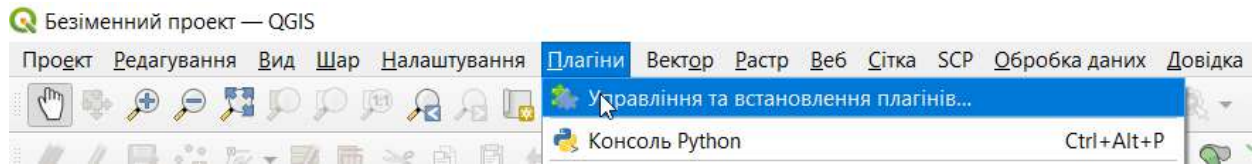
**Мета:** сформувати базові навички роботи з просторовими даними в Q-GIS.

#### Теоретичні питання

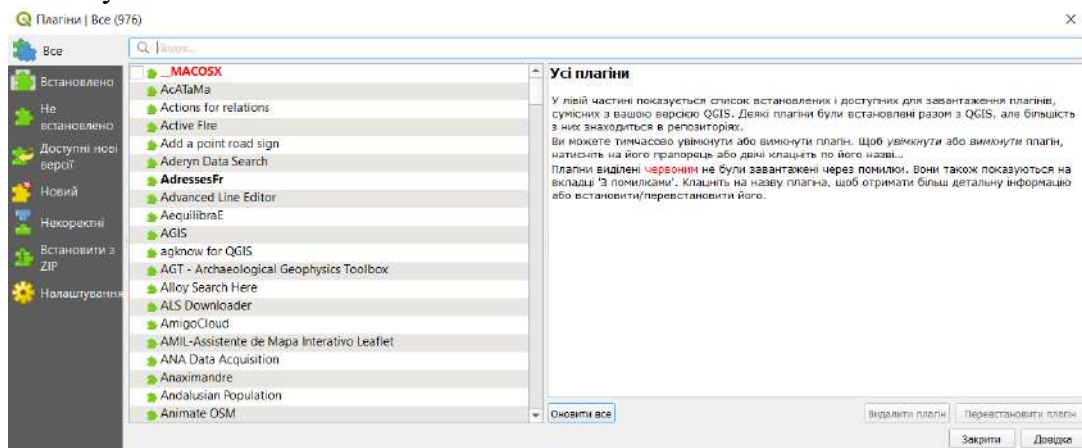
1. Геоінформаційні технології в сучасному світі. Поняття про геоінформаційні системи. Інформатика, геоінформатика, геоінформаційні технології і географія.
2. Визначення ГІС. Відмінність ГІС від інших інформаційних систем.
3. Історія розвитку геоінформаційних технологій.
4. Функції й галузі застосування ГІС і геоінформаційних технологій.
5. Класифікація сучасних ГІС.

#### Хід роботи

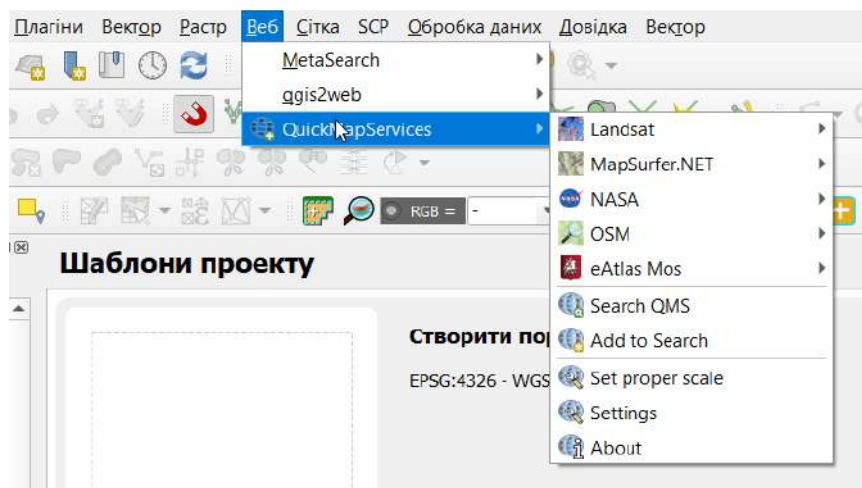
1. Картографічна основа проекту. Встановлення та використання плагіна «QuickMapServices». Переходимо у вкладку «Плагіни» – «Керування та встановлення модулів».



У діалоговому вікні обираємо «Не встановлено» і в рядку пошуку задаємо «QuickMapServices». Натискаємо кнопку «Встановити плагін».

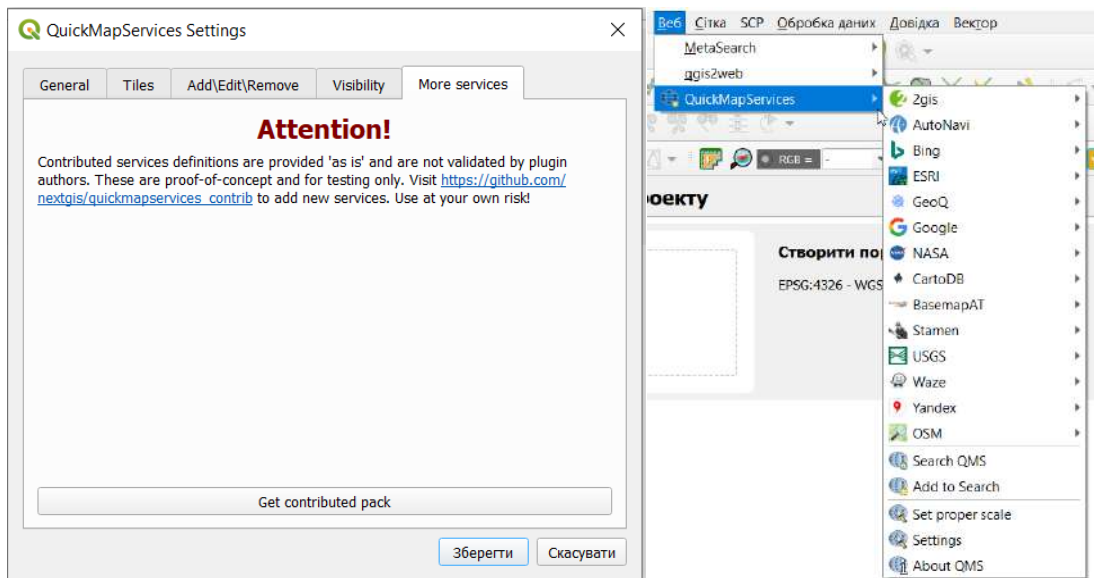


Встановлений плагін буде доступний у меню «Веб».

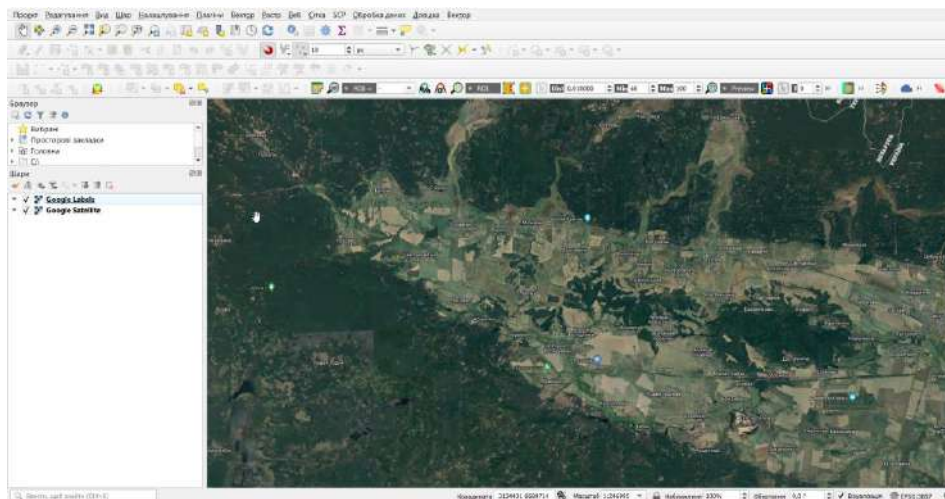


Початково буде доступний дуже обмежений набір карт. Для розширення цього переліку потрібно змінити налаштування. Перейдіть до розділу «Settings» в меню плагіна.

У діалоговому вікні потрібно обрати вкладку «More services», а в ній – «Get contributed pack». Після завантаження карт натисніть «Зберегти». У результаті цього меню плагіна суттєво розшириться.



Спробуйте завантажити різні карти у робочу область зверніть увагу на доступні супутникові зображення поверхні Землі. Для кращого орієнтування за супутниковим зображенням можна завантажити шар «Google Labels».

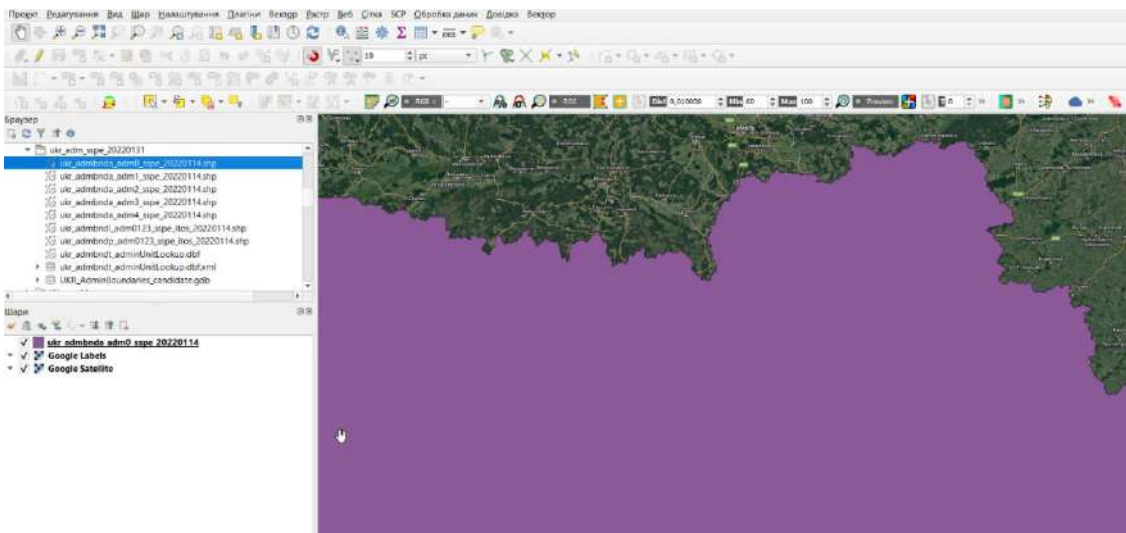


## 2. Додавання та форматування векторних шарів.

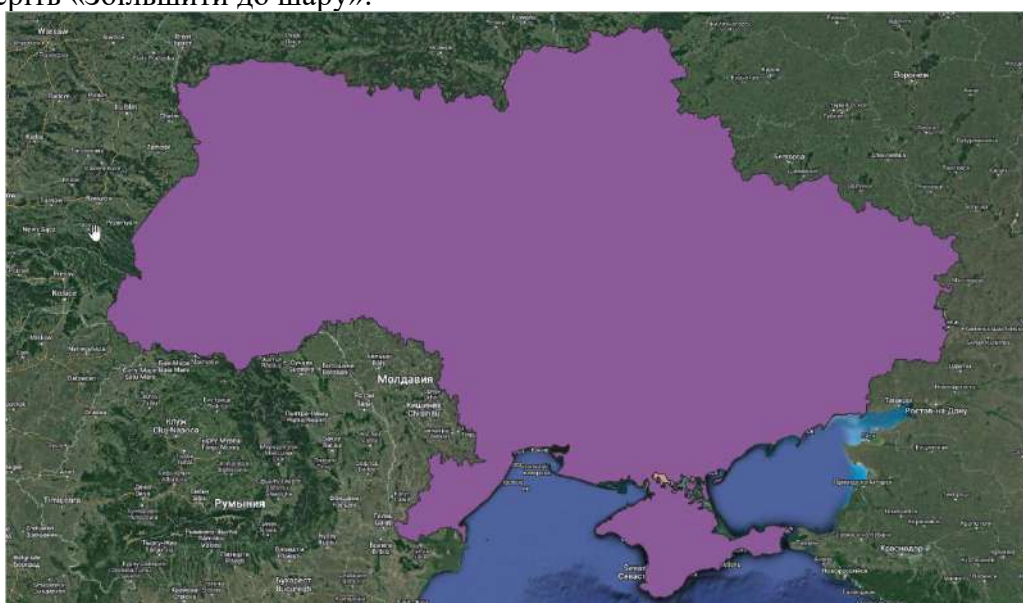
Завантажте і розархівуйте базові векторні шари для території України:  
[https://drive.google.com/file/d/1dWkBiEqN8\\_diuLFJvHK\\_irfJg8-1qKDD/view?usp=share\\_link](https://drive.google.com/file/d/1dWkBiEqN8_diuLFJvHK_irfJg8-1qKDD/view?usp=share_link)


Щоб додати їх на карту скористаємося вікном браузера в інтерфейсі програми Q-GIS. Відшукайте розташування завантаженої папки з шарами на вашому комп'ютері і розкрийте її.

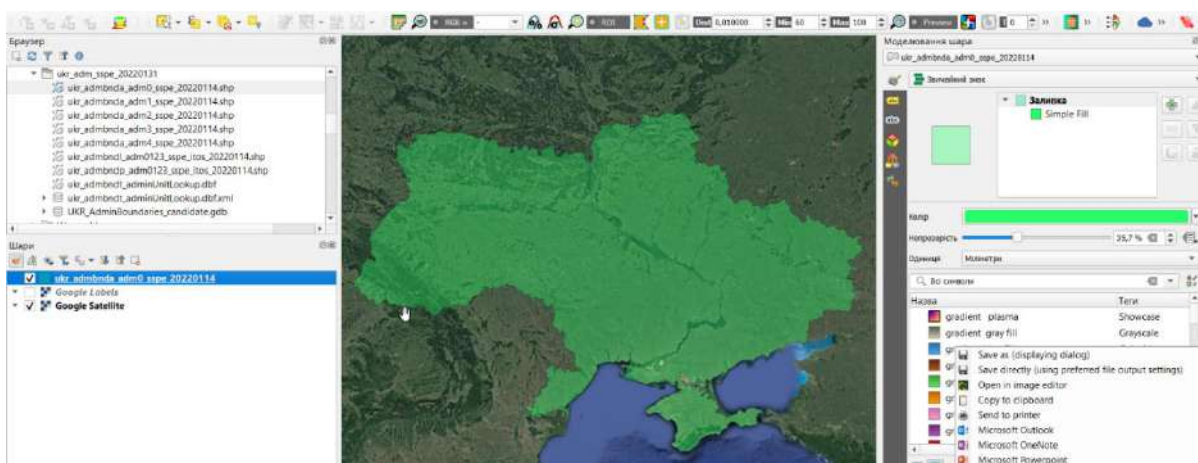
Перетягніть мишкою перший шар ukr\_admbnda\_adm0\_sspe\_20220114 у вікно карти. Новий шар з'явиться у вікні шарів та відобразиться на карті.



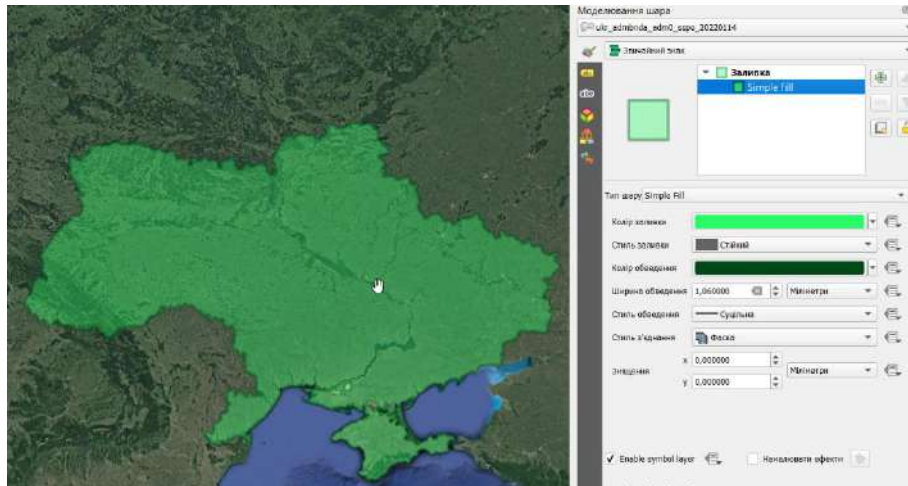
Для відображення повного охоплення шару клікніть по ньому у вікні шарів правою кнопкою мишки та оберіть «Збільшити до шару».



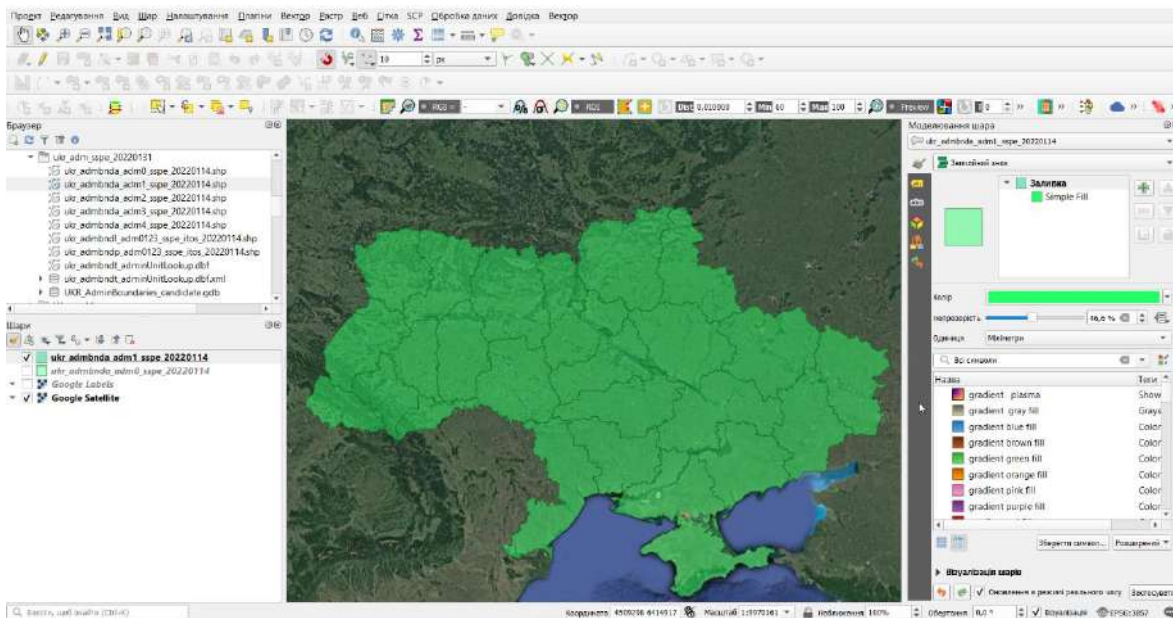
Ми можемо змінювати візуалізацію шару за допомогою меню стилізації шару. Для переходу в це меню скористаємося відповідною піктограмою  вверху вікна шарів. Змінимо колір відображення шару та прозорість.



Також можна змінити колір відображення контуру об'єкта. Для цього клікніть лівою кнопкою мишки по значку стилю шару і змініть колір обведення на темнозелений.



Додайте інші векторні шари для території України та змініть їх стилізацію. Неактуальні шари можна відключити, знявши мітку біля відповідного шару у вікні шарів.

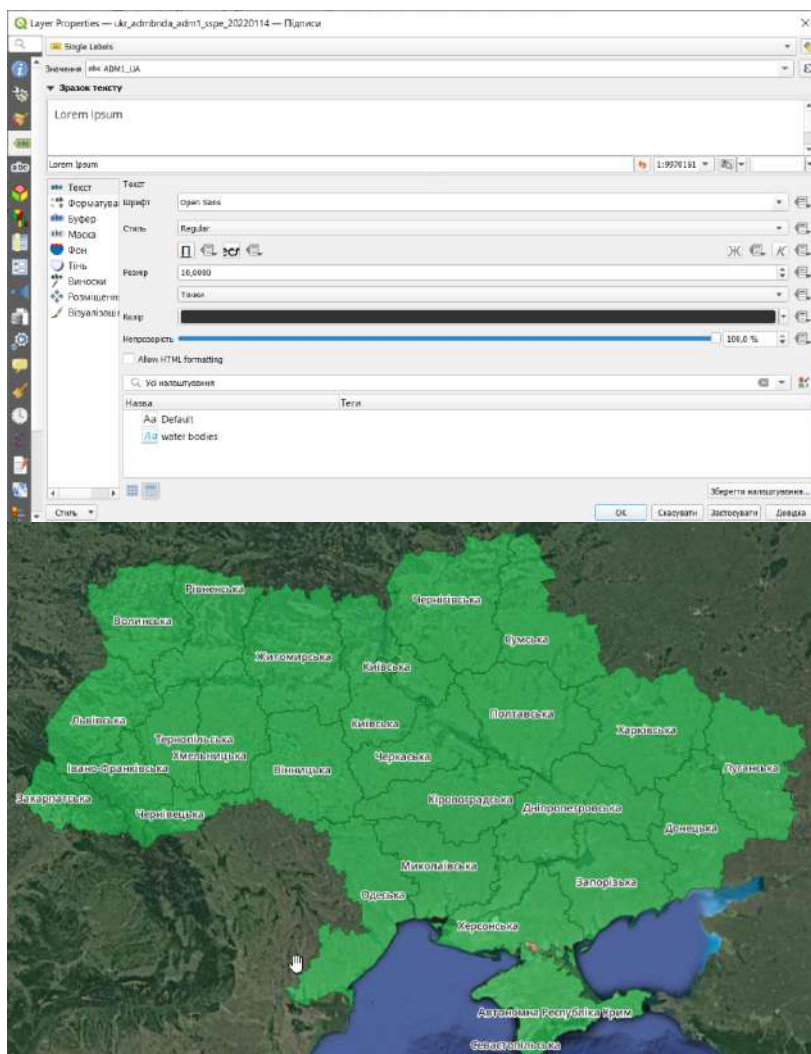


Крім власне просторової інформації векторний шар може містити атрибутивну інформацію, яка відображається у вигляді таблиці атрибутів. Щоб її відкрити, натисніть на шар у вікні шарів правою кнопкою мишки і перейдіть у меню «Відкрити таблицю атрибутів». Як бачимо, таблиця містить додаткову інформацію, прив'язану до просторових об'єктів, відображених у цьому шарі. В даному конкретному випадку це інформація про площу і довжину периметру кожної області України, назви областей трьома мовами та ін. Цю інформацію ми можемо використати для візуалізації шару.

Shape_Leng	Shape_Area	ADM1_EN	ADM1_UA	ADM1_RU	ADM1_PCODE	ADM1_REF	ADM1ALT1EN	ADM1ALT2EN	ADM1ALT1UA	ADM1ALT2UA	ADM1ALT1RU	ADM1ALT2RU	ADM0_EN	ADM0_UA
25,52233518020	2,93150986202	Автономна Рес...	Автономна Рес...	Автономная Р...	UA01	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
12,32636103090	3,25042412030	Vinnitska	Вінницька	Винницкая	UA05	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
12,44077245760	2,59078212324	Volyntsa	Волинська	Волянская	UA07	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
17,96141737490	3,86818425109	Dnipropetrovska	Дніпропетровська	Днепропетровс...	UA12	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
14,78724619510	3,19660466041	Donetska	Донецька	Донецкая	UA14	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
14,11693725570	3,75198175321	Zhytomyrska	Житомирська	Житомирская	UA18	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
8,54274772199	1,54993864828	Zakarpatska	Закарпатська	Закарпатская	UA21	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна
14,09206106280	3,2306086799	Zaporizka	Запорізька	Запорожская	UA23	NULL	NULL	NULL	NULL	NULL	NULL	NULL	Ukraine	Україна

Спочатку підпишемо назви областей. Для цього двічі клікніть мишкою по шару у вікні шарів, перейдіть у вкладку налаштування підписів. Оберіть спосіб відображення міток «Single Labels» та поле таблиці атрибутів ADM1\_UA з українськими назвами областей. Натисніть «Застосувати» та

«ОК». За необхідності можна змінити формат та стиль відображення підписів, використовуючи меню діалогового вікна.



Атрибутивну інформацію ми також можемо використати для стилізації шару. Наприклад, класифікувати області України за їх площею і відобразити їх у вигляді картограми. Для цього перейдіть у меню стилізації шару. Оберіть «Градуирований знак» у якості типу візуалізації та у полі «Значення» оберіть поле таблиці атрибутів «Shape Area». Задайте кольоровий градієнт Reds та натисніть кнопку класифікувати. Області України відобразяться у різних відтінках червоного в залежності від їх площі.



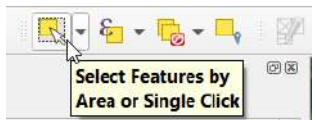
Для збереження результатів роботи і поточного форматування шарів використаємо меню «Проект». Перейдіть у вкладку «Проект» та оберіть «Зберегти як...». Задайте папку на комп'ютері, де ви хочете зберегти проект та його назву (латиницею). Зверніть увагу, що файл проекту має розширення .qgz. Цей файл не містить геоданих, а лише несе інформацію про набір шарів проекту, їх

розташування та оформлення. Якщо окремі шари будуть переміщені після створення проекту – вони відобразяться в проекті не будуть.

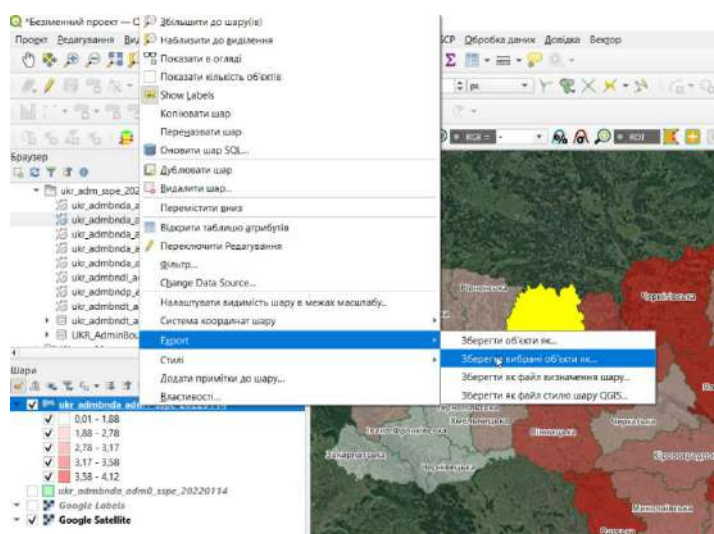
### 3. Створення векторного шару на основі виділення.

Часто виникає необхідність виділити окремий об'єкт з існуючого векторного шару і винести його в окремий шар. Реалізуємо це завдання на прикладі Житомирської області.

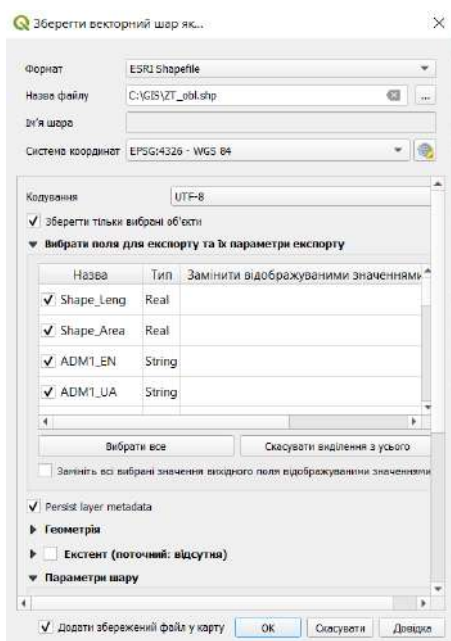
Оберіть інструмент виділення форм на панелі інструментів і клікніть мишкою по Житомирській області.



Житомирська область буде виділена жовтим кольором. Клікніть правою кнопкою мишки по шару і перейдіть у меню «Експорт» – «Зберегти виділені об'єкти як...». У діалоговому вікні оберіть формат файлу Esri shapefile та задайте назву і місце збереження файлів через меню «...» в правій частині рядка «Назва файлу». Натисніть «ОК».

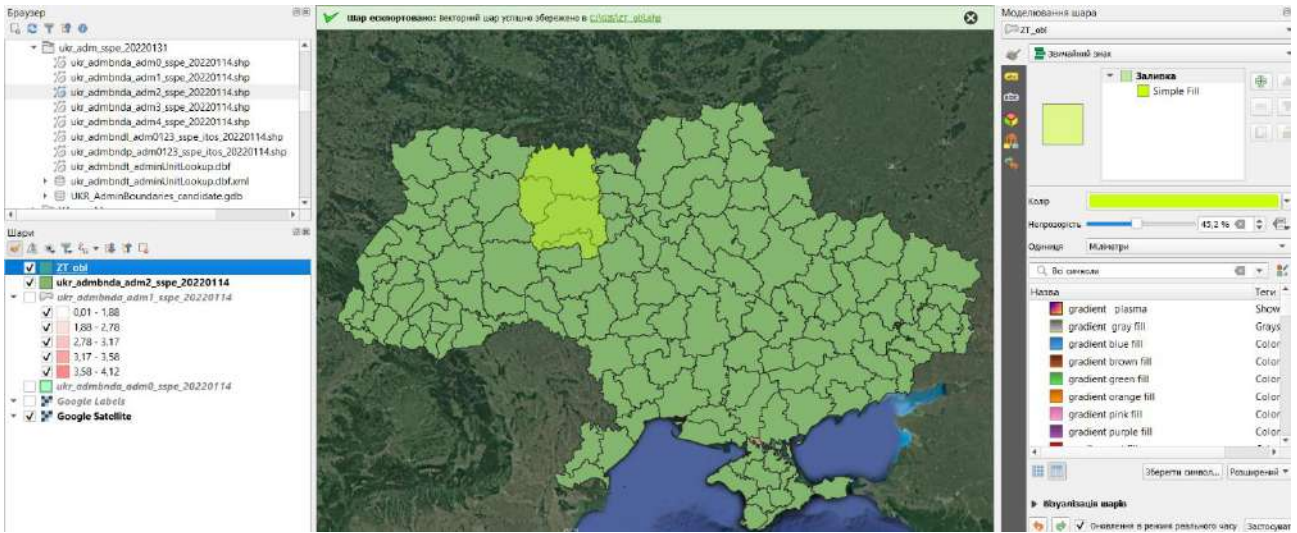


В списку шарів з'явиться новий шар. Попередній шар можете відключити.

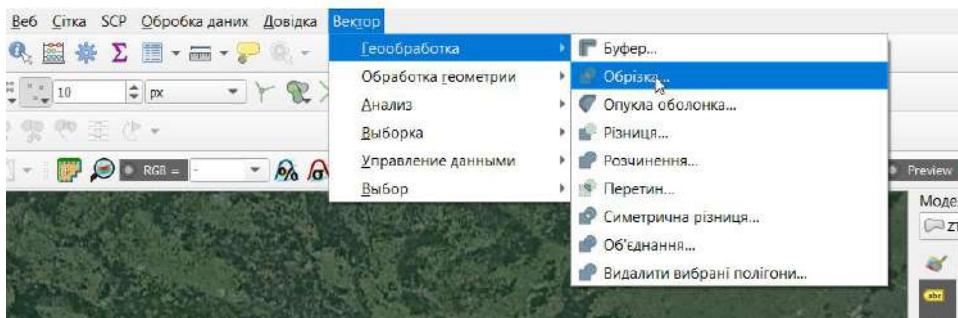


### 4. Обрізання векторного шару за маскою.

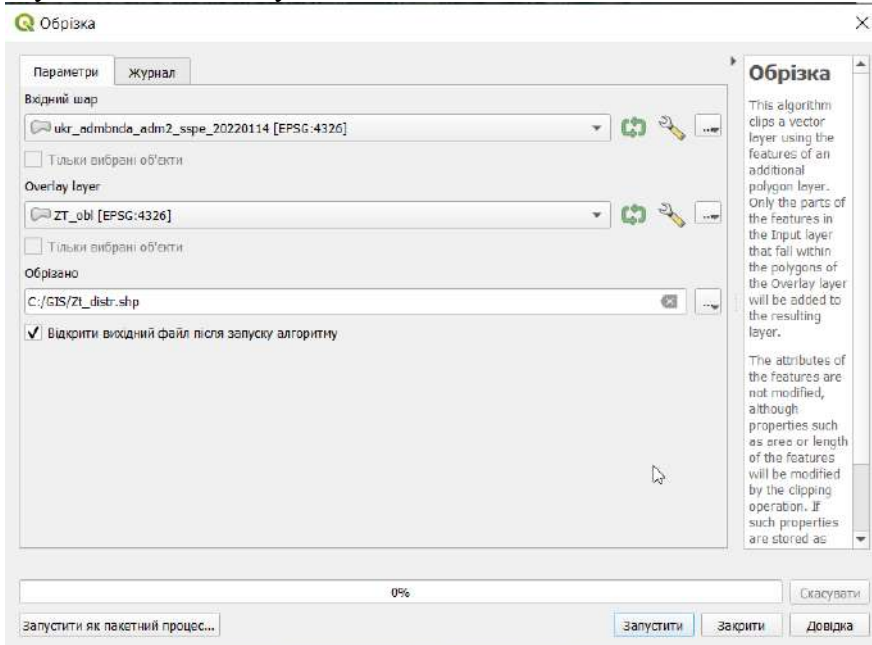
Додамо у вікно карти шар з адміністративним поділом областей на райони.



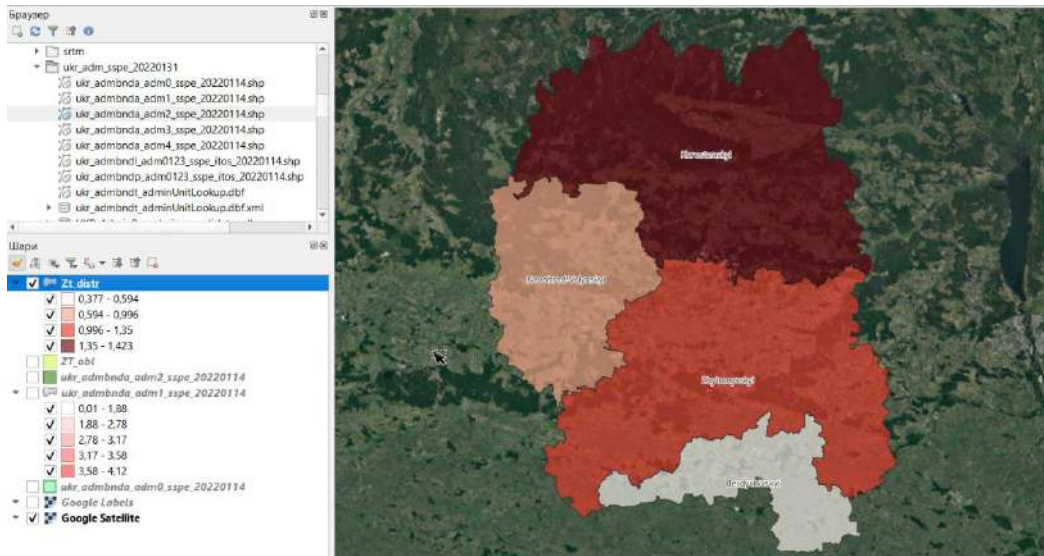
Якщо нас цікавлять лише райони Житомирської області, ми можемо обрізати цей шар за маскою. Перейдіть у меню «Вектор» - «Геообробка» - «Обрізка».



У діалоговому вікні у якості вхідного шару оберіть шар, який потрібно обрізати, а у якості Overlay layer – шар-маску Житомирської області. У рядку «Обрізано» через меню «...» задайте назву та шлях до збереженого файлу. Натисніть «Запустити».

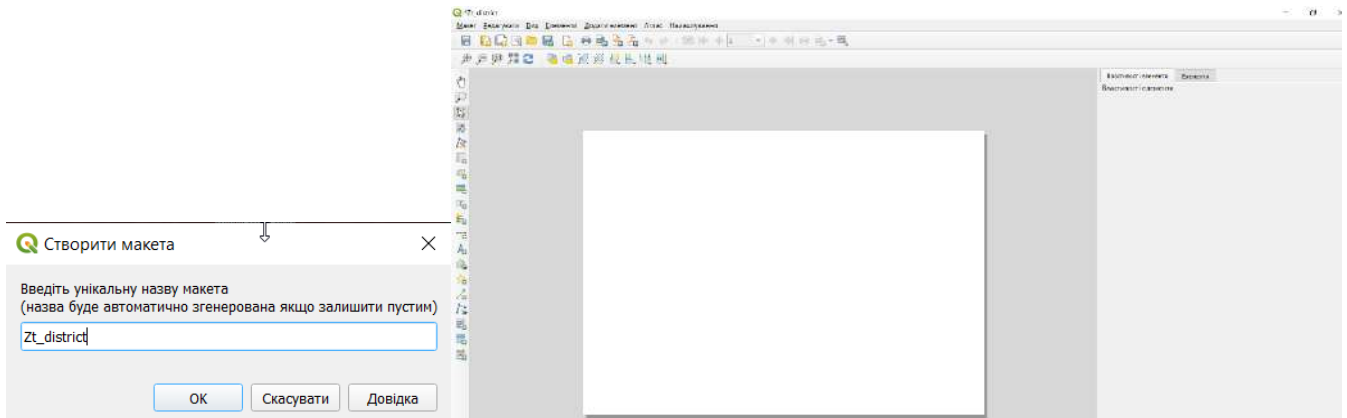


Налаштуйте візуалізацію шару, як показано на малюнку.

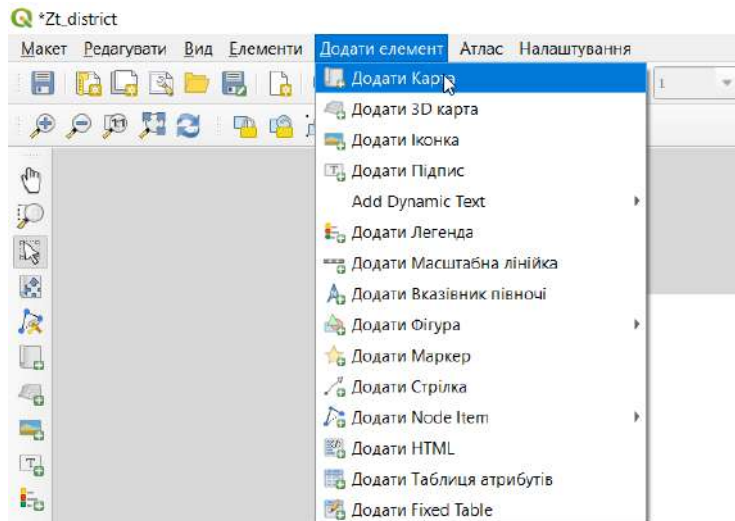


## 5. Підготовка карти до друку.

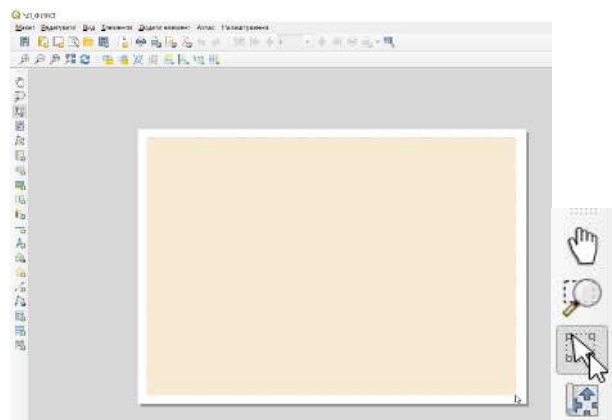
Важливим етапом роботи з картою є створення її макету та підготовка до друку. Для цього в QGIS використовується меню «Макет». Перейдіть в меню «Проект» – «Створити макет друку» і введіть назву макету карти, натисніть «ОК». Відкриється нове вікно з шаблоном проекту, у який буде потрібно додати необхідні елементи карти.



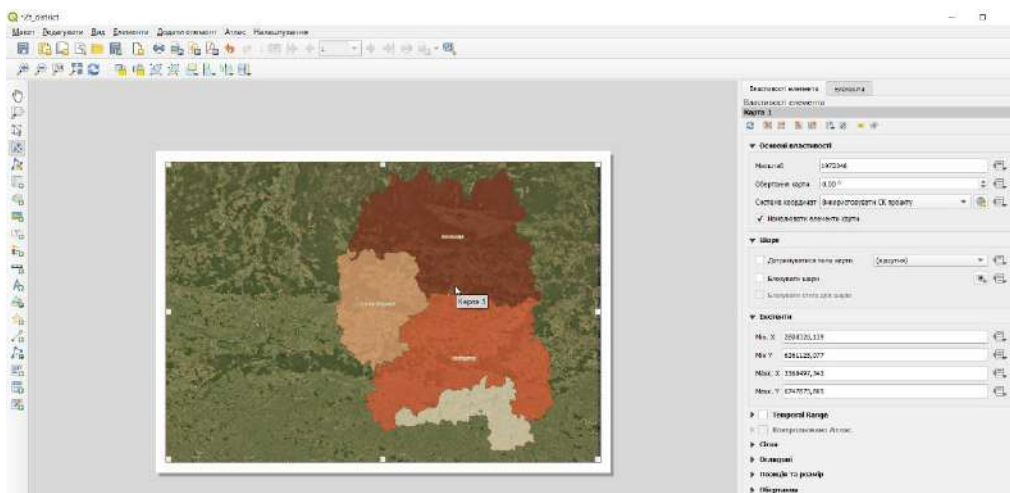
Спочатку додамо картографічне зображення. Перейдіть в меню «Додати елемент» і оберіть «Додати Карта». Зверніть увагу, що через це меню додаються і інші елементи карти. Також доступ до елементів цього меню можливий через вертикальну панель піктограм у лівій частині вікна.




Виділіть за допомогою мишки область розміщення карти на сторінці макету, залишивши невеликі поля і заждіть, поки карта відобразиться. За допомогою верхніх елементів вертикального меню можна здійснювати ряд корисних маніпуляцій з елементами карти. Спробуйте самостійно, як вони працюють.

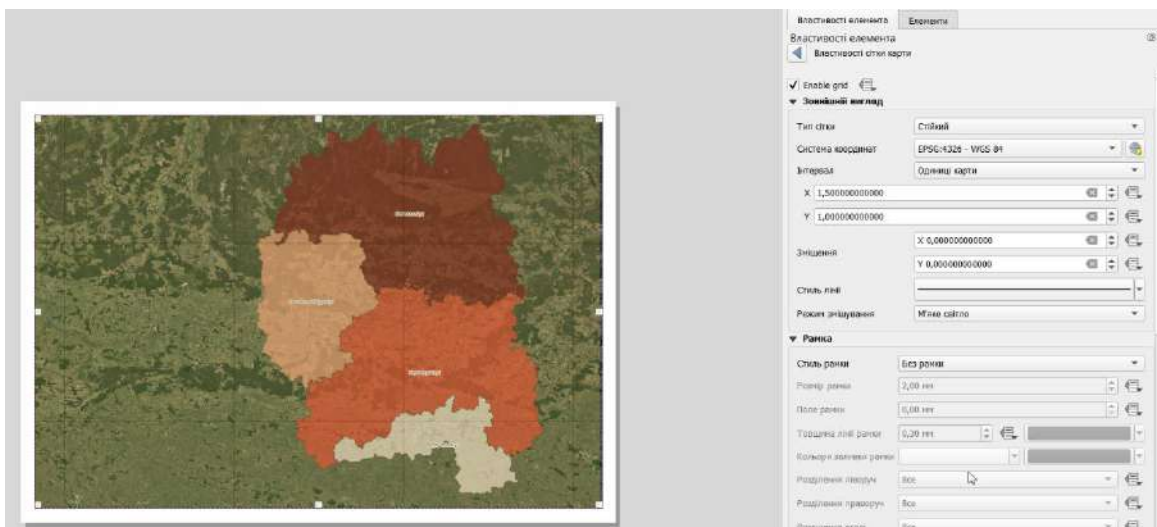


В правій частині вікна є кілька корисних вкладок. За замовчуванням відкриваються «Елементи» – список елементів карти та «Властивості елемента» – відображає властивості поточного елемента. У цій вкладці доступне налаштування візуалізації відповідного елемента.

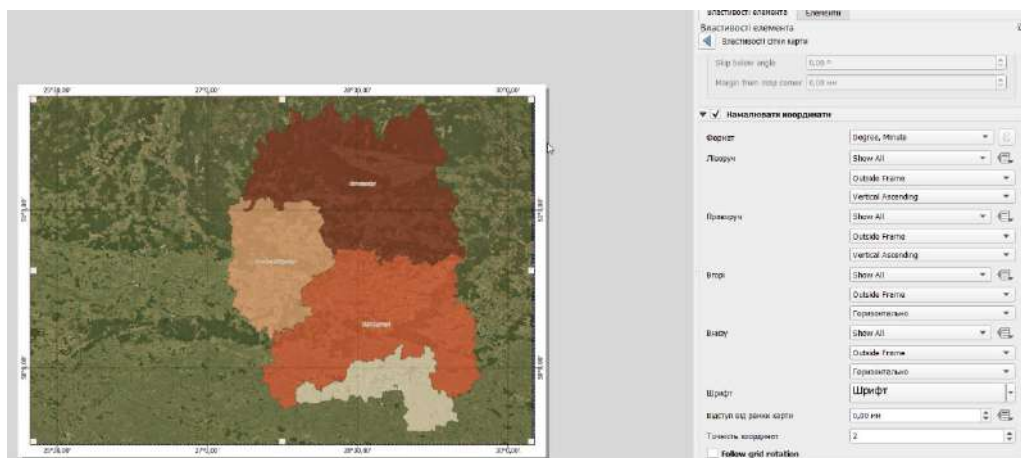


Щоб захистити від змін відформатоване картографічне зображення відмітьте пункти «Блокувати шари» та «Блокувати стилі шарів».

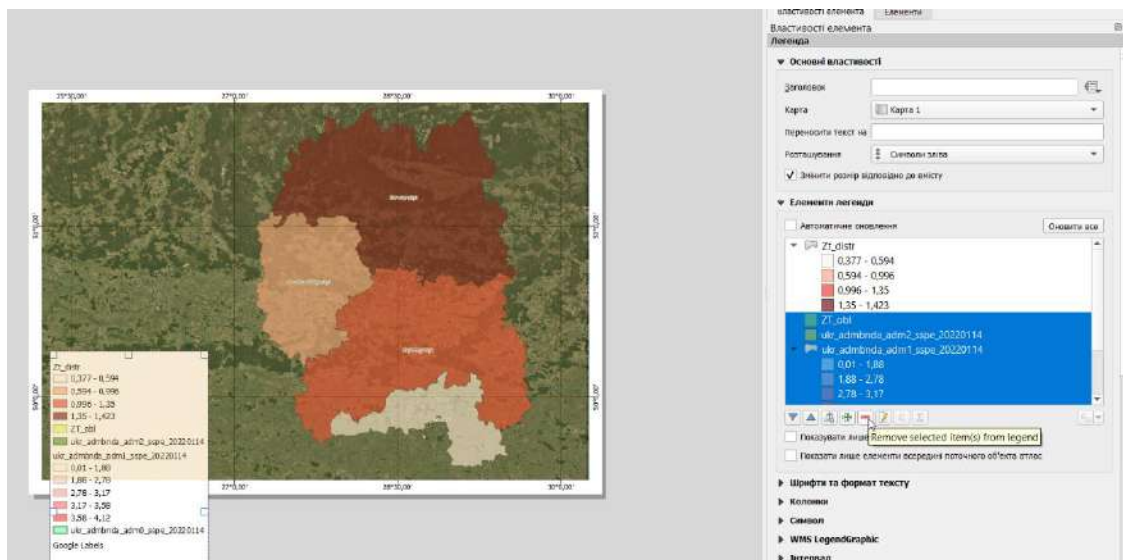
Наступним кроком буде додавання та налаштування координатної сітки. Розгорніть меню «Сітки», додайте нову сітку, натиснувши на  , виділіть створену сітку і натисніть «Змінити...». У вікні властивостей сітки потрібно налаштувати систему координат. Змініть її на WGS 84 і задайте інтервали між лініями сітки. Також можна налаштувати зміщення ліній (для симетричності сітки), стиль ліній та режим змішування. Ці параметри підбираються експериментально в залежності від використаного масштабу і системи координат.



На наступному етапі відобразимо значення координат для ліній сітки. Поставте мітку в меню «Намалювати координати» і налаштуйте їх відображення, як показано на малюнку. Ви можете обрати власний формат координат відповідно до дизайну карти.

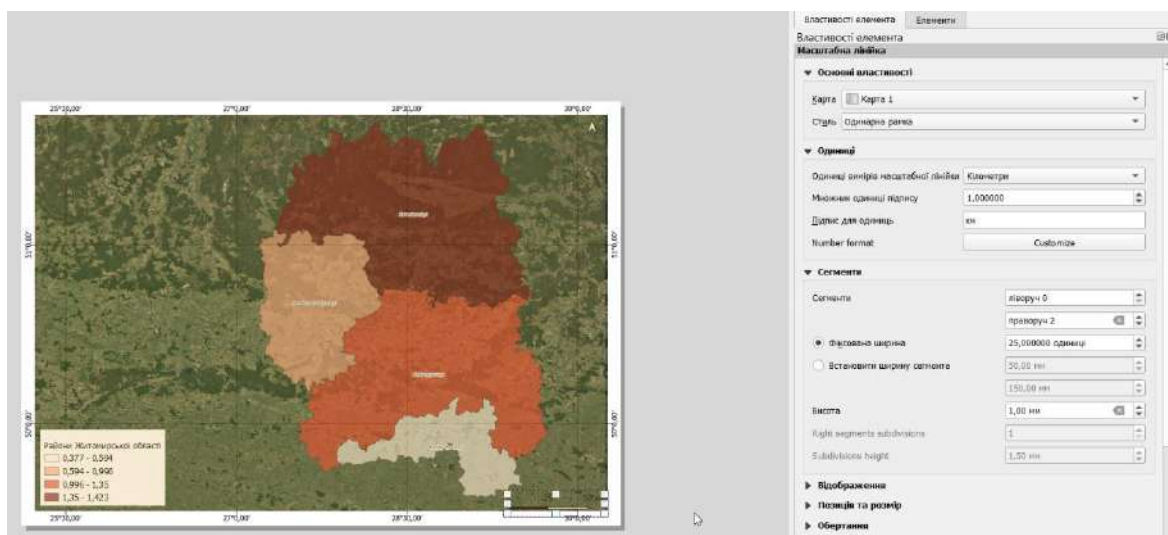


Наступним важливим елементом карти є легенда. Додайте її через відповідне меню. За замовчуванням додається легенда до всіх шарів карти, які є у списку шарів. Щоб видалити зайві елементи легенди відключіть її автооновлення та видаліть зайві елементи.



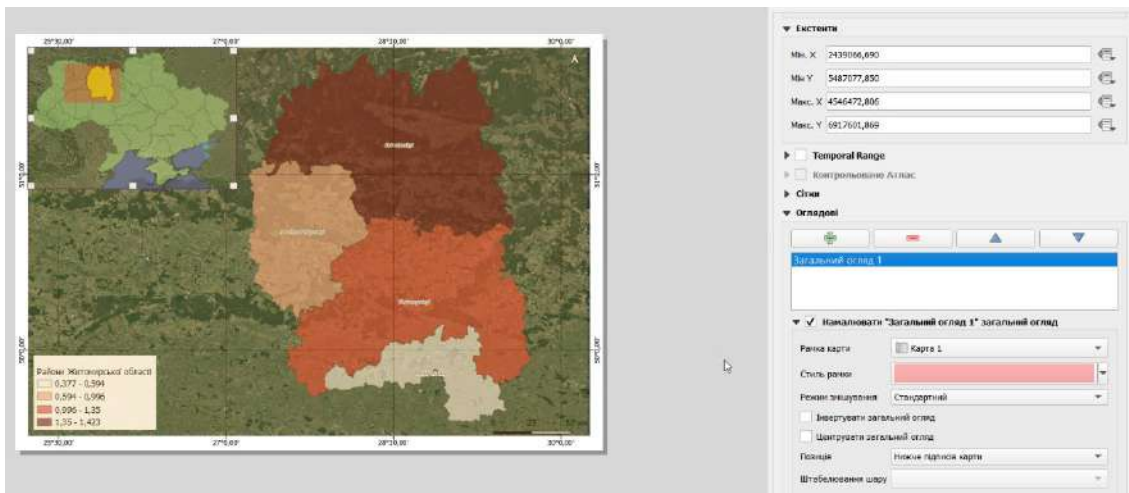
Якщо двічі клікнути мишкою по заголовку легенди – відкриється вікно його редагування. Тут можна змінити цей заголовок.

Додамо до карти масштабну лінійку та вказівник півночі. У меню властивостей цих елементів також можна налаштувати їх візуалізацію.

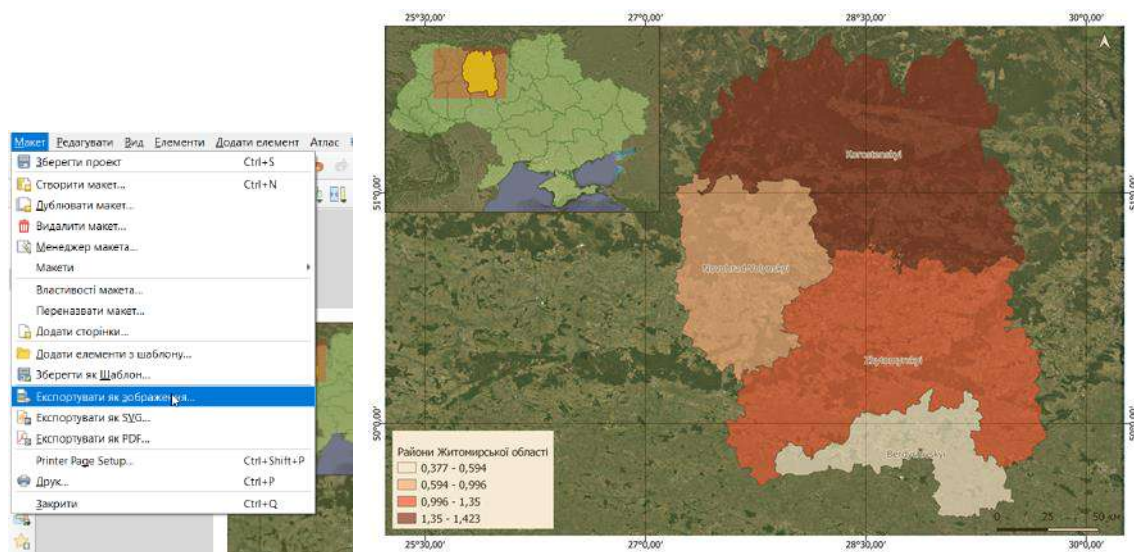


Часто виникає необхідність показати на карті положення регіону інтересу по відношенню до більших одиниць геопростору. Наприклад, положення Житомирської області на карті адміністративного поділу України.

Для цього потрібно підготувати у вікні програми відповідне картографічне зображення і додати ще одну карту в макет. Охоплення основної карти на додатковій можна налаштувати в меню «Оглядові».



Підготовку макету карти завершено. На заключному етапі збережемо його у графічному форматі, придатному для друку, наприклад JPG.



Завантажте створену карту в графічному форматі на гугл-диск для оцінювання.

### Рекомендована література

1. ГІС в екологічних дослідженнях та природоохоронній справі : навч. Посібник / Ю. М. Андрейчук, Т. С. Ямелинець – Львів : «Простір-М», 2015. – 284с. Режим доступу: [https://wwfeu.awsassets.panda.org/downloads/wwf\\_.pdf](https://wwfeu.awsassets.panda.org/downloads/wwf_.pdf)
2. Застосування ГІС у природоохоронній справі на прикладі відкритої програми QGIS: навч. посіб. / О. Часковський, Ю. Андрейчук, Т. Ямелинець. – Львів : ЛНУ ім. Івана Франка, Вид-во Простір-М, 2021. – 228 с.
3. Пітак І.В., Негадайлов А.А., Масікевич Ю.Г., Пляцук Л.Д., Шапорев В.П., Моїсєєв В.Ф. Геоінформаційні технології в екології : Навчальний посібник. – Чернівці, 2012.– 273с. Режим доступу: <https://essuir.sumdu.edu.ua/bitstream/123456789>

## Лабораторна робота № 7

**Тема:** Приєднання атрибутів за значенням поля таблиці. Візуалізація атрибутивних даних.

**Мета:** навчитись візуалізувати статистичні дані (атрибути) картографічними засобами

### Теоретичні питання

1. Апаратне забезпечення геоінформаційних систем і технологій.
2. Загальна характеристика апаратного забезпечення ГІС.
3. Пристрої збору і введення інформації.
4. Пристрої візуалізації і подання даних.
5. Тенденції розвитку апаратного забезпечення.

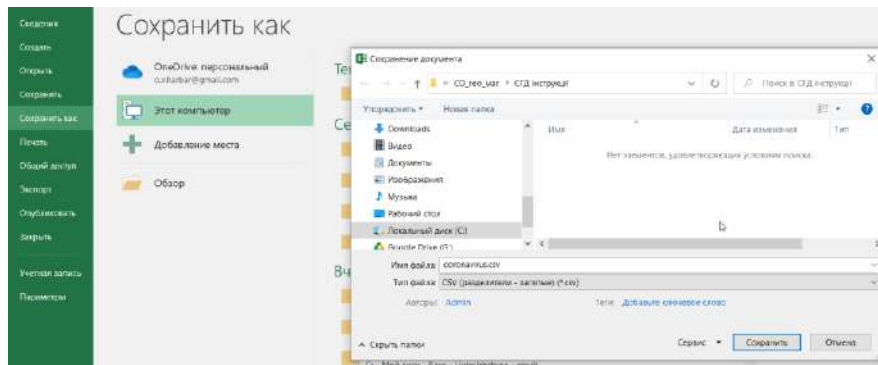
### Хід роботи

#### 1. Приєднання атрибутів за значенням поля таблиці.

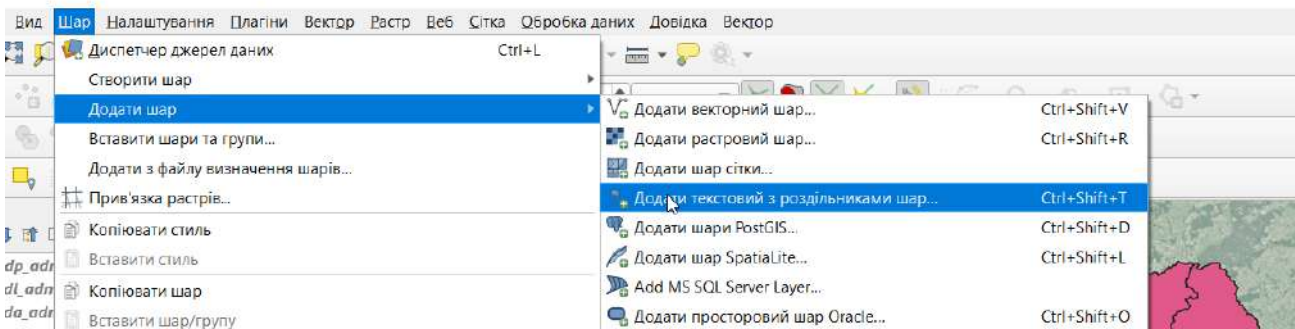
Для побудови картограми можна використати не лише внутрішні атрибути бази даних. Але і зовнішні, приєднавши їх до існуючої таблиці атрибутів за спільним полем. Використаємо дані медичної статистики щодо захворюваності на Covid-19 по областях України, наведені в таблиці.

N	Region	Detected	Died	Recovered
1	Vinnytska	145319	3458	119637
2	Volynska	124838	2342	107704
3	Dnipropetrovska	284714	9289	241314
4	Donetska	203090	5463	173919
5	Zhytomyrska	183470	3467	151432
6	Zakarpatska	109000	2657	89497
7	Zaporizka	204152	5738	180807
8	Ivano-Frankivska	177904	3407	143187
9	Kyiv	411576	8770	350397
10	Kyivska	216161	4912	190429
11	Kirovohradska	41282	1703	33413
12	Luhanska	74678	2333	60637
13	Lvivska	295051	6437	226009
14	Mykolaiivska	135956	3455	117885
15	Odeska	308595	5810	236071
16	Poltavska	168078	4084	146276
17	Rivnenska	165353	2482	142335
18	Sumska	174474	3120	143845
19	Ternopilska	141911	2345	114876
20	Kharkivska	293771	6743	247227
21	Khersonska	100167	2833	79513
22	Khmelnyska	192131	3608	157972
23	Cherkaska	153885	2877	135625
24	Chernivetska	147608	3240	120733
25	Chernihivska	119253	2682	100890

Скопіюйте таблицю і вставте на лист Excel у першу клітинку та збережіть цей лист у форматі csv:

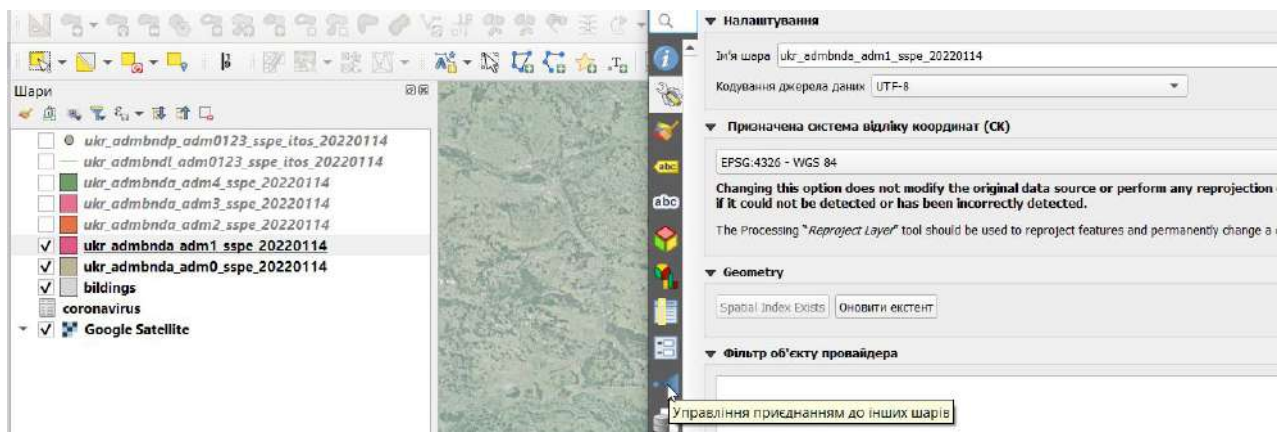


Додайте створений документ у проект Q-gis через меню «Шар»:

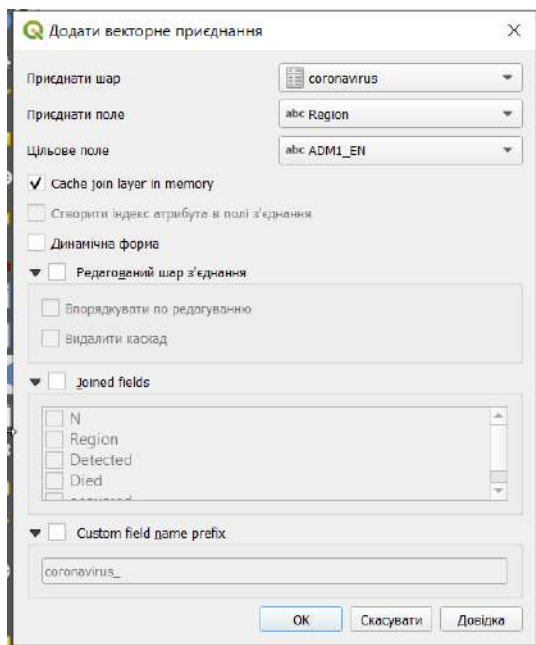


Ми будемо приєднувати дані до шару з полігонами областей України за їх назвами. У створеній таблиці це поле «Область», а у шарі областей (ukr\_admbnda\_adm1\_sspe\_20220114) – це поле ADM\_1EN.

Клікніть мишкою двічі по шару і у вікні налаштування оберіть опцію «Управління приєднанням до інших шарів».

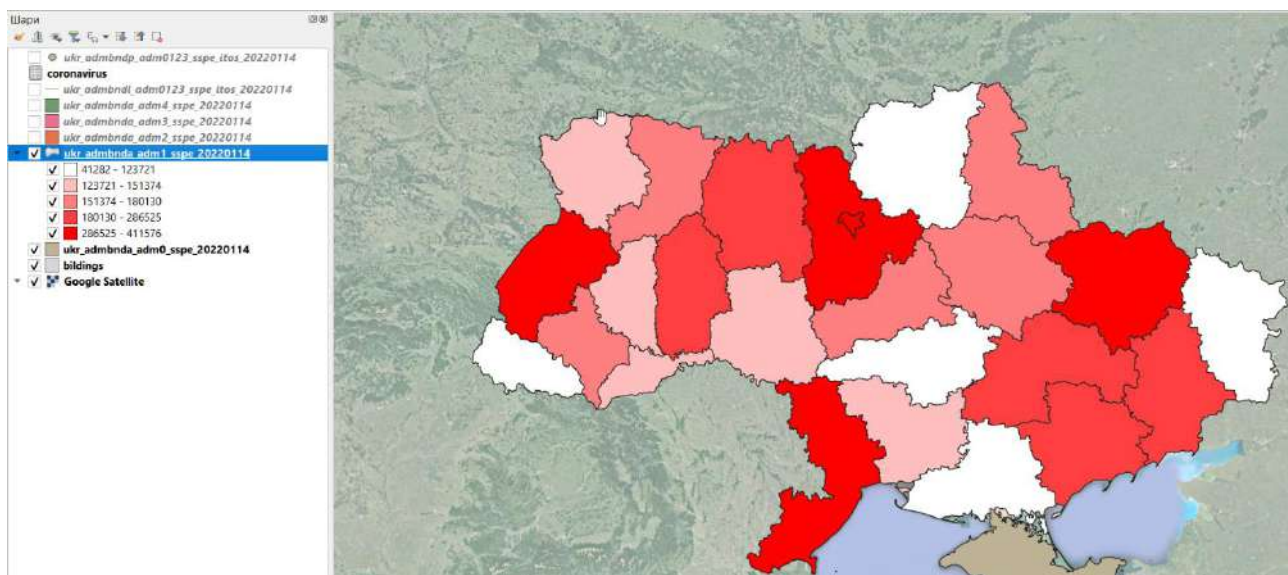


Натисніть на кнопку «Додати новий зв'язок» (+) і у діалоговому вікні задайте налаштування, як на скриншоті і натисніть «Ок» та «Застосувати» і «Ок» у наступному вікні. Відкрийте таблицю атрибутів шару ukr\_admbnda\_adm1\_sspe\_20220114 і переконайтесь, що дані приєднались.



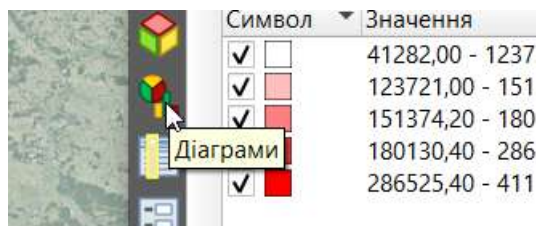
validTo	coronavirus_N	coronavirus_Detected	coronavirus_Died	coronavirus_recovered
NULL	NULL	NULL	NULL	NULL
NULL	23	153 865	2 877	135 625
NULL	25	119 253	2 682	100 890
NULL	24	147 608	3 240	120 733
NULL	3	284 714	9 289	241 314
NULL	4	203 090	5 463	173 919
NULL	8	177 904	3 407	143 187
NULL	20	293 771	6 743	247 227
NULL	21	100 167	2 833	79 513
NULL	22	192 131	3 608	157 972
NULL	11	41 282	1 703	33 413
NULL	9	411 576	8 770	350 397
NULL	9	411 576	8 770	350 397
NULL	12	74 678	2 333	60 637

Побудуйте картограму з використання даних поля coronavirus\_Detected – число виявлених випадків захворювання. Підготуйте макет карти до друку, як описано в першій роботі, і завантажте на форум для оцінювання.

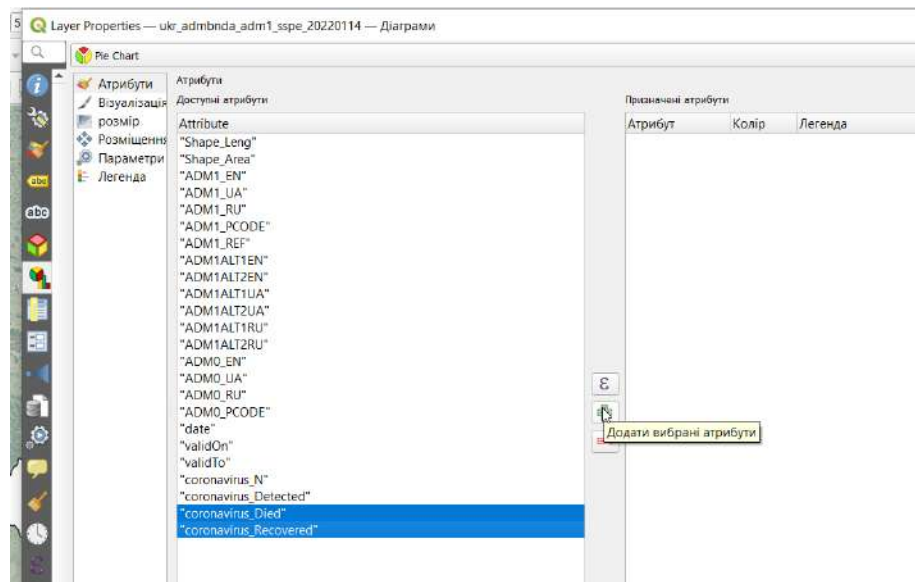


## 2. Побудова секторальних картодіаграм.

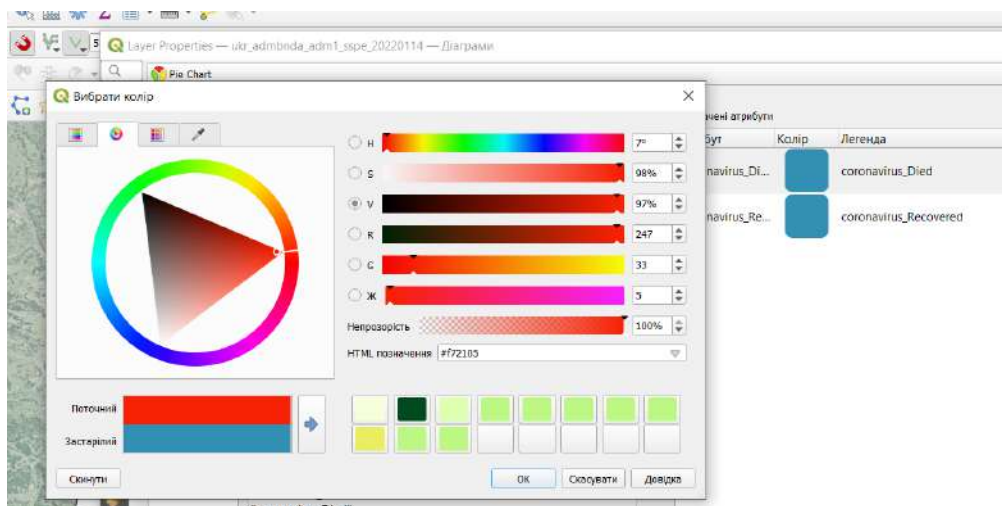
Дані двох інших полів coronavirus\_Died та coronavirus\_Recovered використаємо для побудови картодіаграм. Подвійним кліком по шару перейдіть до меню налаштувань шару і оберіть опцію «Діаграми»



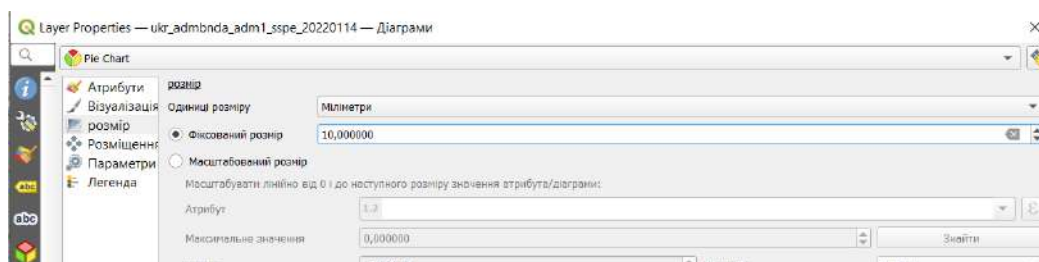
Оберіть тип діаграми Pie Chart та атрибути для побудови діаграм:

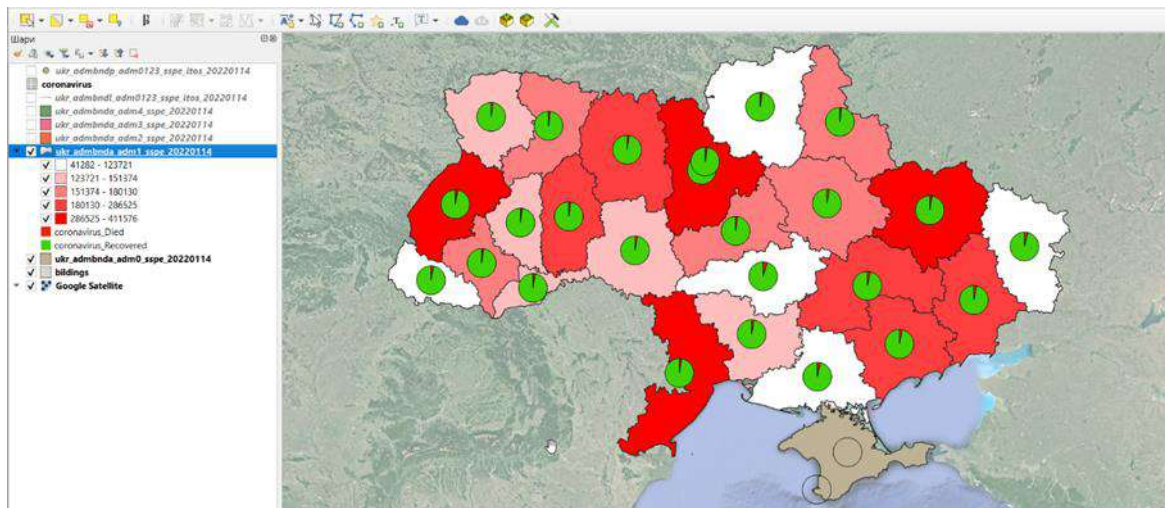


Подвійним кліком по полю «Колір» відкрийте діалог налаштування кольорів і налаштуйте їх для кожного атрибута.



Налаштуйте розмір діаграм і натисніть «Ок».



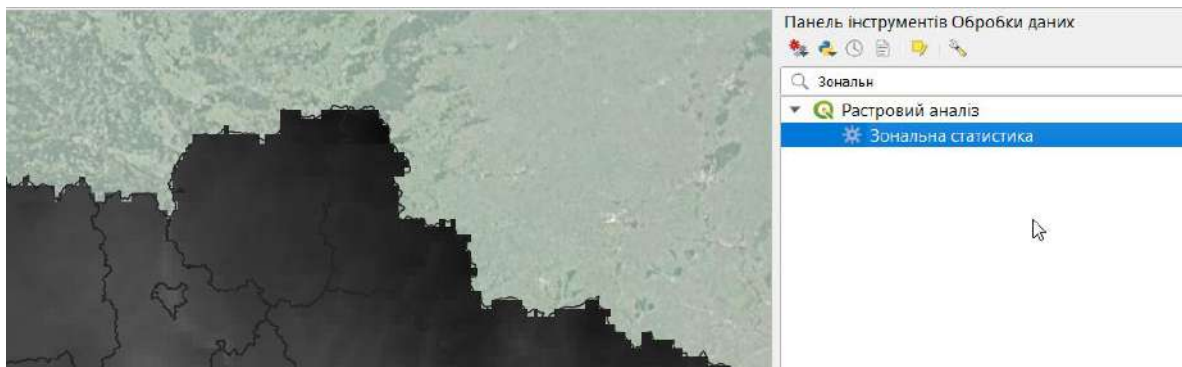


**Підготуйте макет карти до друку. Завантажте створену карту на гугл-диск для оцінювання.**

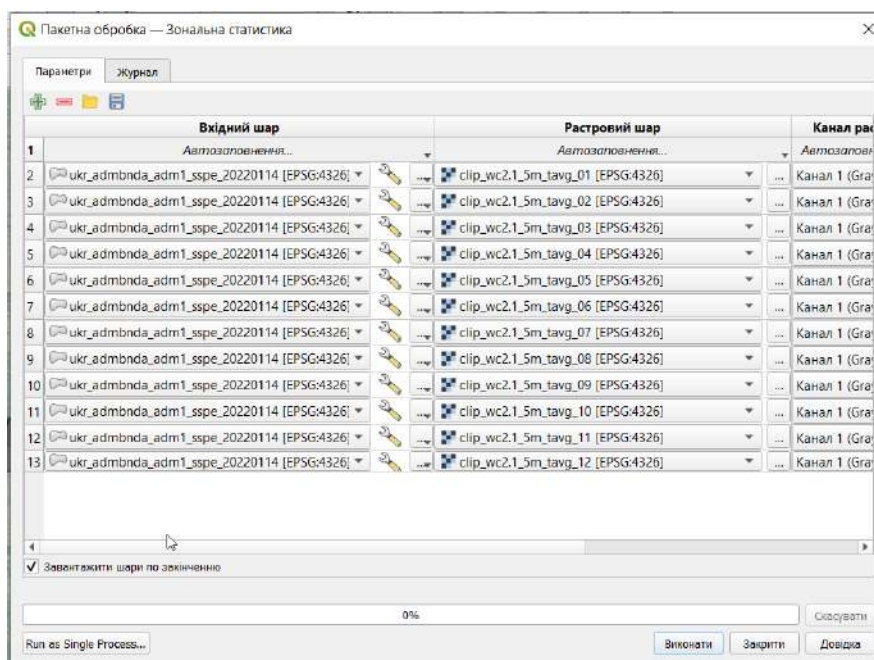
### Рекомендована література

1. ГІС в екологічних дослідженнях та природоохоронній справі : навч. Посібник / Ю. М. Андрейчук, Т. С. Ямелинець – Львів : «Простір-М», 2015. – 284с. Режим доступу: [https://wwfeu.awsassets.panda.org/downloads/wwf\\_.pdf](https://wwfeu.awsassets.panda.org/downloads/wwf_.pdf)
2. Застосування ГІС у природоохоронній справі на прикладі відкритої програми QGIS: навч. посіб. / О. Часковський, Ю. Андрейчук, Т. Ямелинець. — Львів : ЛНУ ім. Івана Франка, Вид-во Простір-М, 2021. – 228 с.
3. Пітак І.В., Негадайлов А.А., Масікевич Ю.Г., Пляцук Л.Д., Шапорев В.П., Моїсеєв В.Ф. Геоінформаційні технології в екології : Навчальний посібник. – Чернівці, 2012.– 273с. Режим доступу: <https://essuir.sumdu.edu.ua/bitstream/123456789>

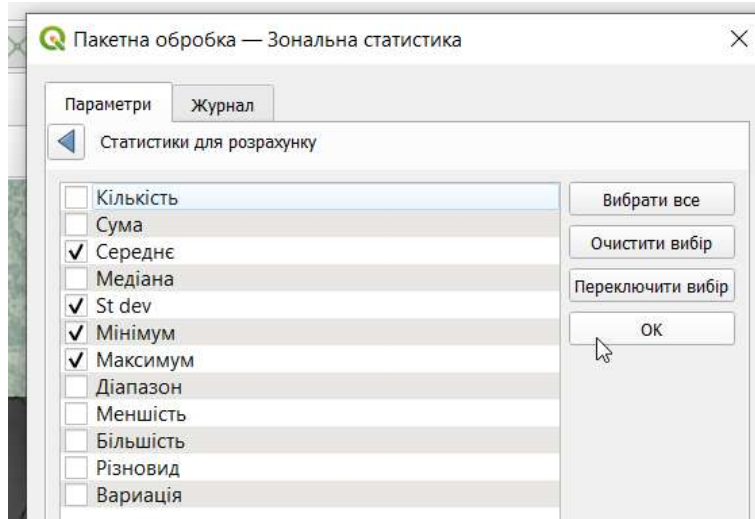




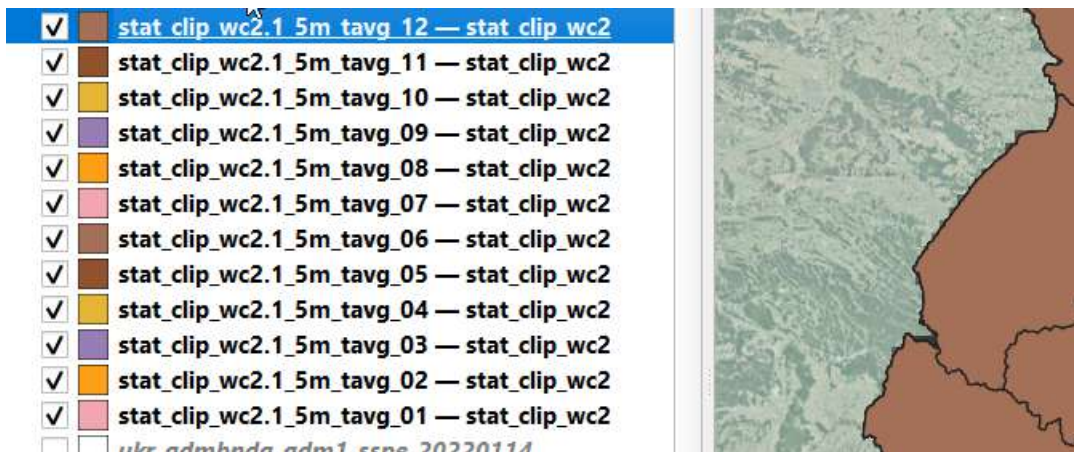
Задайте вхідний шар `ukr_admbnda_adm1_sspe_20220114`. Оскільки ми збираємось отримати помісячні показники, запустимо процес як пакетний. Задайте вихідні шари, як показано на рис. і поставте галочку біля опції «Завантажити шари по закінченню».



У колонці «Статистики для розрахунку» оберіть «Середнє», «Мінімум» і «Максимум» і натисніть «ОК». Автоматично заповніть поле для інших шарів.



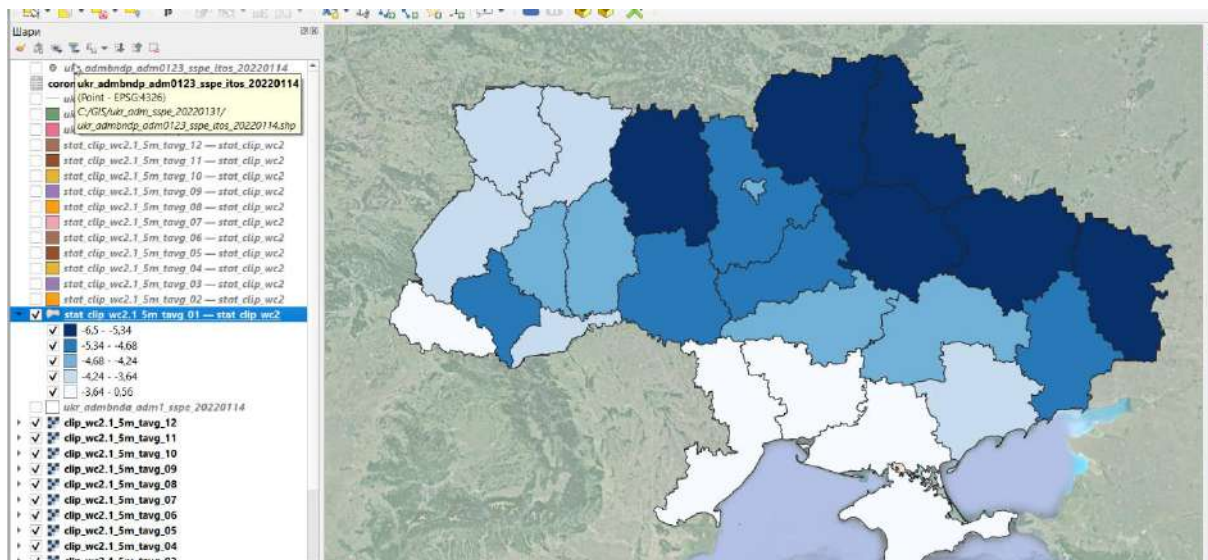
У наступному вікні потрібно задати шлях та назви файлів для збереження статистичних показників. Використаємо формат геопакетів та автозаповнення назвами растрових шарів, додавши до них «stat». У основному вікні програми після завершення процесу з'явиться 12 нових полігональних шарів, відповідно місяцям року.



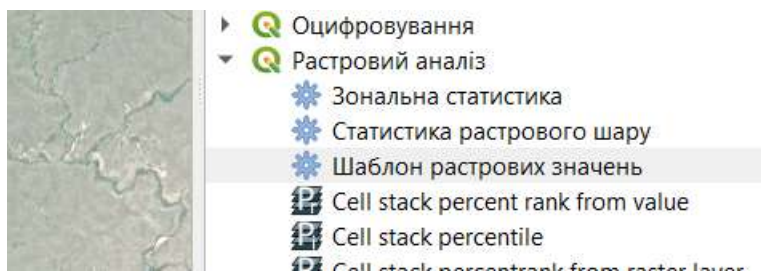
Якщо відкрити таблиці атрибутів цих шарів, можна побачити дані, згенеровані інструментом зональної статистики – середнє арифметичне, мінімум і максимум для кожної області по використаному кліматичному показнику.

date	validOn	validTo	coronavirus_N	coronavirus_Detecte	coronavirus_Died	coronavirus_Recover	_mean	_min	_max
09.11.21	14.01.22	NULL	NULL	NULL	NULL	NULL	0,10655220939...	-3,09500002861...	3,42100000381...
09.11.21	14.01.22	NULL	1	145319	3458	119637	-4,90400216041...	-6,60400009155...	-2,83500003814...
09.11.21	14.01.22	NULL	2	124838	2342	107704	-3,85233421350...	-4,38600015640...	-3,34700012207...
09.11.21	14.01.22	NULL	3	284714	9289	241314	-4,54312613418...	-5,58500003814...	-2,90899991989...
09.11.21	14.01.22	NULL	4	203090	5463	173919	-5,13954392999...	-6,35500001907...	-3,28860759735...
09.11.21	14.01.22	NULL	5	183470	3467	151432	-5,58660514088...	-6,58099985122...	-4,40199995040...

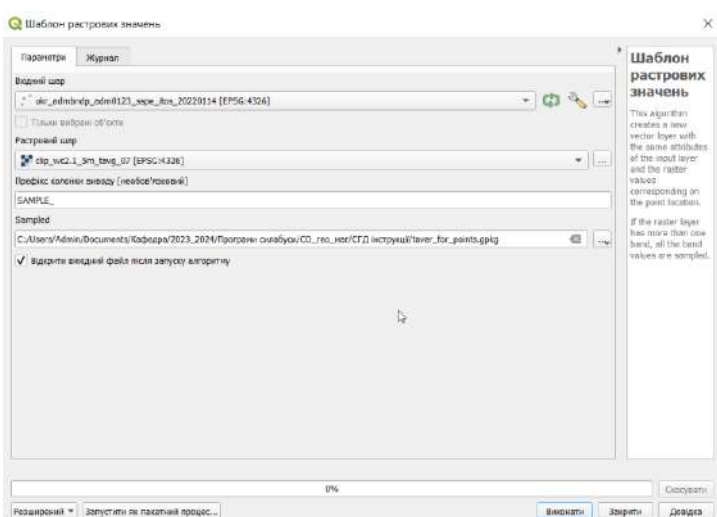
Використаємо середні значення для візуалізації шару та підготовки макету карти.



Ми можемо також розрахувати статистики для заданих точок. У якості таких точок можна використати центроїди територіальних громад з шару ukr\_admbndp\_adm0123\_sspe\_itos\_20220114. Перейдіть до інструментів растрового аналізу і оберіть «Шаблон растрових значень».



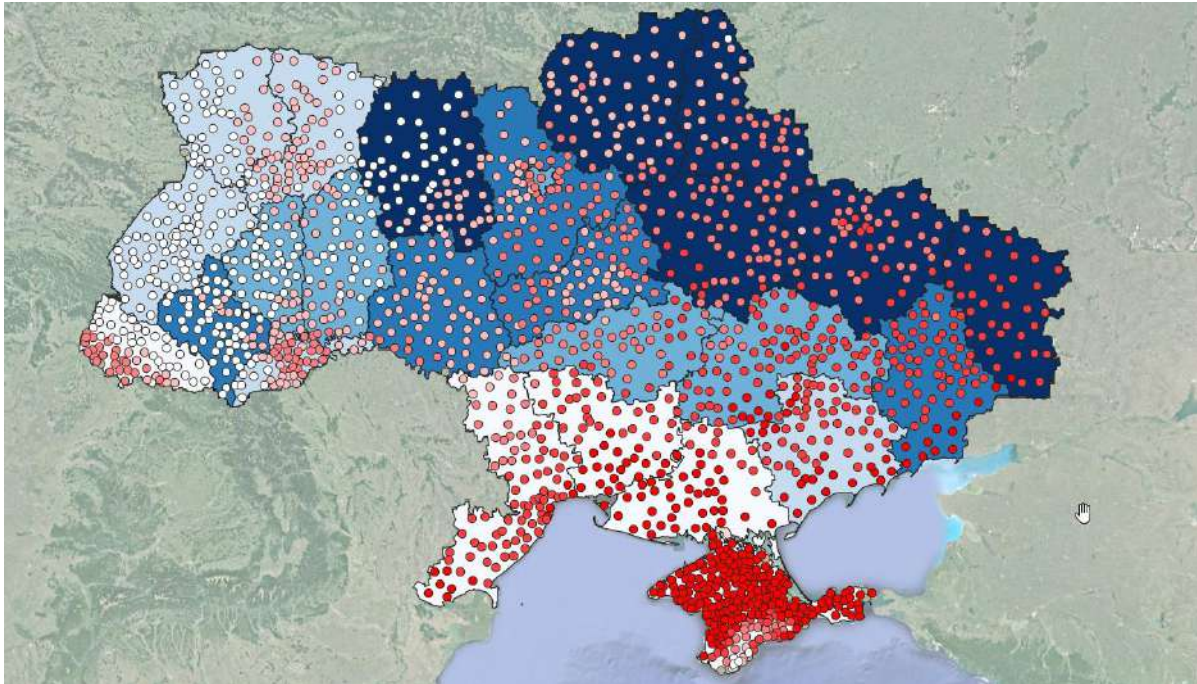
У якості вхідного шару задайте ukr\_admbndp\_adm0123\_sspe\_itos\_20220114, а у якості растрового шару оберіть середні температури липня (stat\_clip\_wc2.1\_5m\_tavg\_07 — stat\_clip\_wc2). Задайте шлях збереження та назву файлу результатів та натисніть «Виконати».



У списку шарів з'явиться новий точковий шар. Відкрийте його таблицю атрибутів. В кінці таблиці ви побачите новий параметр SAMPLE\_1, який містить значення пікселів растру у заданих точках. Тобто середню температуру липня в центроїді кожної територіальної громади.

validTo	POINT_X	POINT_Y	SAMPLE_1
NULL	33,6684293065	44,5757234918	21,5610008239...
NULL	33,8252163628	44,8048143104	22,3290004730...
NULL	33.8873886012	44.7372454992	21.5919990539...

Цей параметр можна використати для стилізації точкового шару.



В результаті роботи ми отримали точковий шар, який демонструє розподіл по території України значень середніх температур липня та полігональний шар, який показує розподіл середніх температур січня по областях України.

**Підготуйте макет карти, використовуючи ці шари. Завантажте карту на гугл-диск для оцінювання.**

#### Рекомендована література

1. ГІС в екологічних дослідженнях та природоохоронній справі : навч. Посібник / Ю. М. Андрейчук, Т. С. Ямелинець – Львів : «Простір-М», 2015. – 284с. Режим доступу: [https://wwfeu.awsassets.panda.org/downloads/wwf\\_.pdf](https://wwfeu.awsassets.panda.org/downloads/wwf_.pdf)
2. Застосування ГІС у природоохоронній справі на прикладі відкритої програми QGIS: навч. посіб. / О. Часковський, Ю. Андрейчук, Т. Ямелинець. – Львів : ЛНУ ім. Івана Франка, Вид-во Простір-М, 2021. – 228 с.
3. Пітак І.В., Негадайлов А.А., Масікевич Ю.Г., Пляцук Л.Д., Шапорев В.П., Моїсєєв В.Ф. Геоінформаційні технології в екології : Навчальний посібник. – Чернівці, 2012.– 273с. Режим доступу: <https://essuir.sumdu.edu.ua/bitstream/123456789>

## Лабораторна робота №9

### Тема: Картування оселища інвазивного виду *Opuntia humifusa* (Cactaceae) на основі біокліматичних даних та алгоритму MAXENT

**Мета:** навчитися проводити біокліматичне моделювання екологічної ніші біологічного виду.

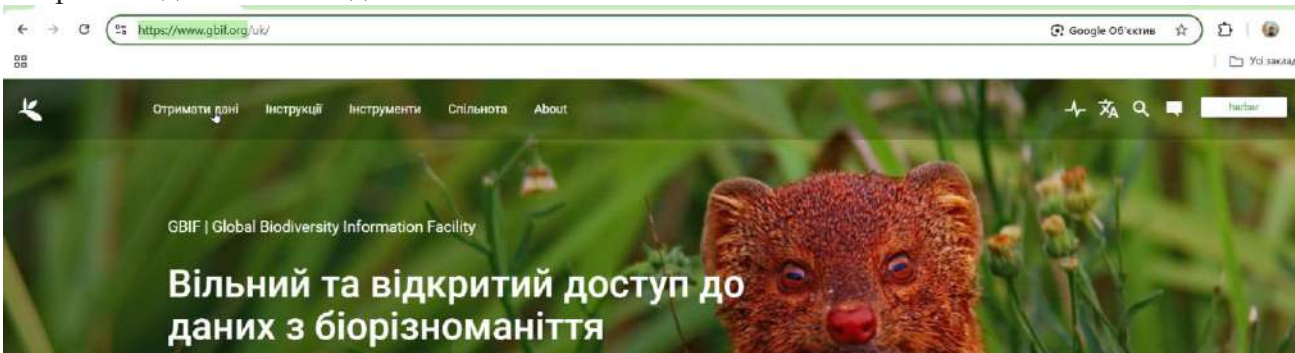
#### Теоретичні питання

1. Використання ГІС в біологічних дослідженнях.
2. Поняття екологічної ніші та кореляційний підхід до прогнозування поширення видів.
3. Загальний алгоритм побудови корелятивних моделей екологічної ніші.
4. Основні завдання, які вирішуються методом моделювання екологічних ніш.

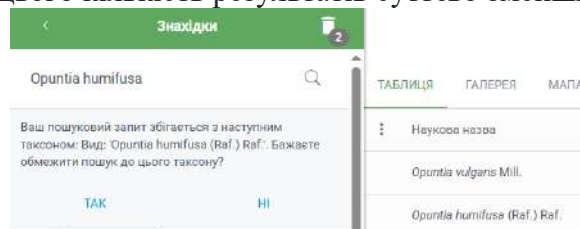
#### Хід роботи

Для виконання цієї роботи необхідні два набори даних: дані про поширення виду – географічні координати спостережень виду та екологічні предиктори - параметри навколишнього середовища у растровому форматі, які містять або безперервні, або категоріальні значення, такі як опади, температура, висота над рівнем моря тощо. У цій роботі ми використаємо дещо спрощений підхід, оскільки вона має на меті лише ознайомитись із загальною методологією картування оселищ видів (моделювання поширення видів) з використанням алгоритму та програмного забезпечення MAXENT. Потрібно розуміти, що на практиці процедура картування оселища виду набагато складніша і потребує широкого залучення експертних оцінок. Окрім загальнодоступних даних про поширення виду бувають необхідними додаткові польові обстеження з метою уточнення даних а екологічні предиктори не обмежуються загальнодоступними кліматичними даними. Часто виникає потреба створювати додаткові набори даних у растровому форматі.

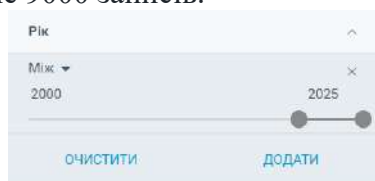
Для цієї роботи ми будемо використовувати дані з Глобальної інформаційної системи з біорізноманіття (GBIF, <https://www.gbif.org>). Після нескладної реєстрації (через Гугл-акаунт) ви можете завантажити дані самостійно. Для цього авторизуйтеся на сайті GBIF та перейдіть у вкладку «Отримати дані» - «Знахідки».



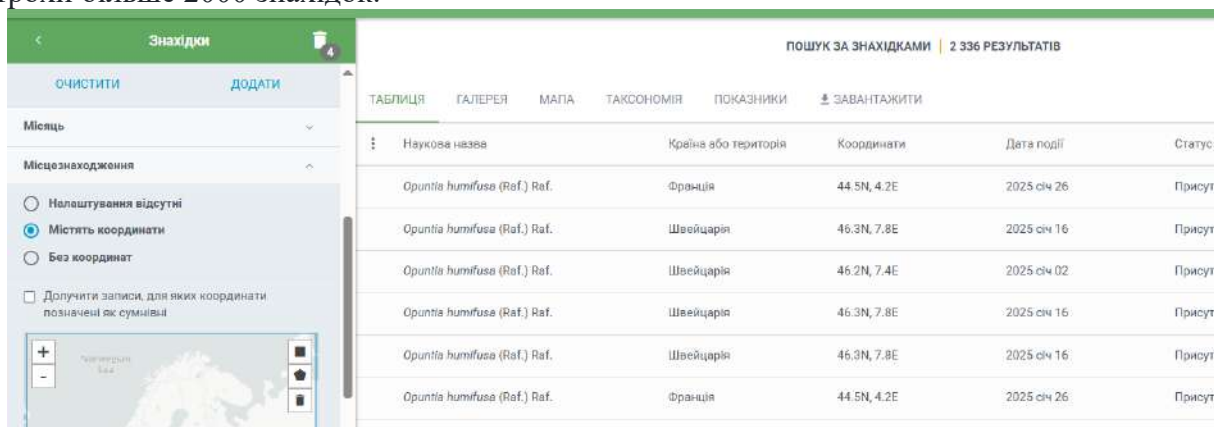
У рядку пошуку введіть назву виду, та запустіть пошук. Ви побачите, що відфільтровано більше ніж 500000 знахідок і програма запитує чи потрібно обмежитись вказаним видом. Вірогідно, початково дані було відібрано з використанням двох ключових слів. Тому погоджуємось обмежитись лише вказаним видом. Після цього кількість результатів суттєво зменшиться (менше 13000).



Також доцільним буде використати часовий фільтр. У базі є дуже старі дані за минулі століття. Нічого не відомо про те, чи збереглися ці популяції. Тому обмежимося лише 21 століттям. Після цього залишиться трохи більше 9000 записів.



Потрібно також пам'ятати, що батьківщина цього виду Північна Америка і він поширився майже по всіх континентах. А ми збираємось моделювати його екологічну нішу лише у Європі. Тому доцільно застосувати ще і просторовий фільтр. У категорії «Місцезнаходження» оберіть пункт «Містять координати» та за допомогою інструменту «Прямокутне виділення» оберіть на карті територію всієї Європи. Висока точність тут не потрібна, однак краще задати трохи більше охоплення. Зайві дані потім можна буде видалити. У наслідок застосування усіх фільтрів залишиться лише трохи більше 2000 знахідок.



Варто також переглянути карту знахідок. Як бачимо, вид широко розповсюджений на півдні Європи. У північному напрямку число реєстрацій різко знижується і це відображає реальне поширення виду, а не похибки реєстрації. В Україні ця закономірність підтверджується.

Натисніть кнопку «Завантажити» і оберіть завантажити «CSV». Через певний час дані будуть згенеровані і стануть доступними для завантаження. Зверніть увагу, що на сторінці завантаження одразу вказано посилання для цитування цих даних (GBIF.org (19 August 2025) GBIF Occurrence Download <https://doi.org/10.15468/dl.xearyf>). З метою економії часу ми використаємо попередньо завантажені дані, які містяться у архіві 0023153-25081113504898.zip (саме в такому вигляді завантажуються дані. Розархівуйте їх. Ви побачите, що архів містить один csv-файл. При спробі відкрити його безпосередньо ви не зможете прочитати дані. Цей файл потрібно відкрити через Excel. Створіть новий файл Excel, назвіть його «*opuntia\_humifusa*» та відкрийте його. Перейдіть у вкладку «Дані» - «Отримання зовнішніх даних» - «Із тексту» та оберіть на комп'ютері файл 0023153-25081113504898.csv.

У діалоговому вікні оберіть «Мої дані містять заголовки» та перейдіть далі. Тут ви можете переглянути зразок даних. Якщо вони відображаються адекватно – перейдіть далі та оберіть текстовий формат відображення даних для полів з назвою виду та географічними координатами (в іншому випадку деякі координати можуть бути розпізнані як дата), натисніть «Готово» та вставте дані на поточний лист. Після цього збережіть дані. Оскільки для подальшої роботи нам потрібні лише три з них: назва виду, географічна широта та географічна довгота, інші поля можна видалити. Варто також змінити назву колонок з координатами так, як вказано на малюнку.

species	Latitude	Longitude	
<i>Opuntia humifusa</i>	42.75652	-9.1114	
<i>Opuntia humifusa</i>	42.25209	-8.80745	
<i>Opuntia humifusa</i>	43.435799	11.838442	
<i>Opuntia humifusa</i>	46.235434	7.364492	

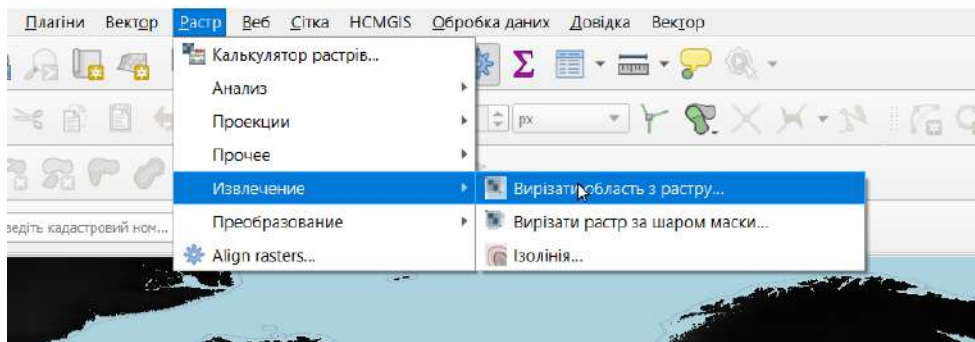
Підготовку даних про поширення майже завершено. Залишилось зберегти дані у csv форматі, який буде зрозумілий для програмного забезпечення. Однак щоб під час збереження даних координати не були розпізнані як дати, потрібно крапки змінити на коми. Після збереження відкрийте файл csv у «Блокнот» та замініть коми на крапки. Після цього замініть розділювач колонок «;» на «,». Збережіть файл.

У якості екологічних предикторів ми будемо використовувати набори растрових даних WorldClim (<https://www.worldclim.org/>). WorldClim — це набір растрових шарів глобальних кліматичних даних, які можна використовувати для картографування та екологічного моделювання. Для цієї роботи ми будемо використовувати біокліматичні змінні (усереднені дані за 1970-2000 pp.)

для сучасних умов. Використаємо дані з роздільною здатністю 2,5 кутових хвилин (~ 4,6 км). Така роздільна здатність буде цілком достатньою для території моделювання. З метою економії часу ви можете скористатись попередньо завантаженими даними з архіву wc2.1\_2.5m\_bio.zip. Розархівуйте його. Ви побачите, що архів містить растрові дані глобального охоплення для 19 біокліматичних змінних (<https://www.worldclim.org/data/bioclim.html>).

Біокліматичні змінні отримують з щомісячних значень температури та кількості опадів для створення більш біологічно значущих змінних. Вони часто використовуються в моделюванні поширення видів та пов'язаних з ним методах екологічного моделювання. Біокліматичні змінні відображають річні тенденції (наприклад, середньорічну температуру, річну кількість опадів), сезонність (наприклад, річний діапазон температури та кількості опадів) та екстремальні або лімітувальні фактори навколишнього середовища (наприклад, температуру найхолоднішого та найтеплішого місяця, а також опади вологих та сухих кварталів).

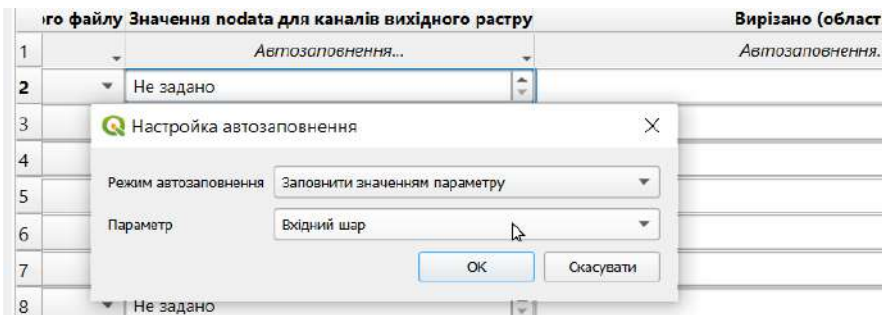
Відкрийте растрові шари біокліматичних змінних та точковий шар з даними про поширення виду в Q-GIS. Нам потрібно обрізати растрові шари за екстентом території дослідження. Перейдіть до інструменту обрізання растру, як показано на малюнку.



Натисніть кнопку «Запустити як пакетний процес», щоб обрізати всі растри одночасно. Додайте ще 18 рядків для вхідних шарів та оберіть у кожному рядку одну із біокліматичних змінних. У полі «Область, що вирізається» натисніть чорний трикутник в кінці рядка і оберіть «намалювати на полотні карти». Після цього протягують мишкою по полотну карти позначте необхідний екстент. Щоб використати його і для інших змінних використайте автозаповнення.



У колонці «Вирізане» натисніть «...» і змініть тип вихідного файлу на ASC, оскільки це тип файлу, який використовує MaxEnt. У полі назви файлу можете поставити 1. Створіть також папку для збереження файлів wc2.1\_2.5m\_bio\_ASC. У наступному вікні оберіть «Заповнити значенням параметру» та «вхідний шар». Це дозволить використати назву вхідних шарів у назвах обрізаних шарів.



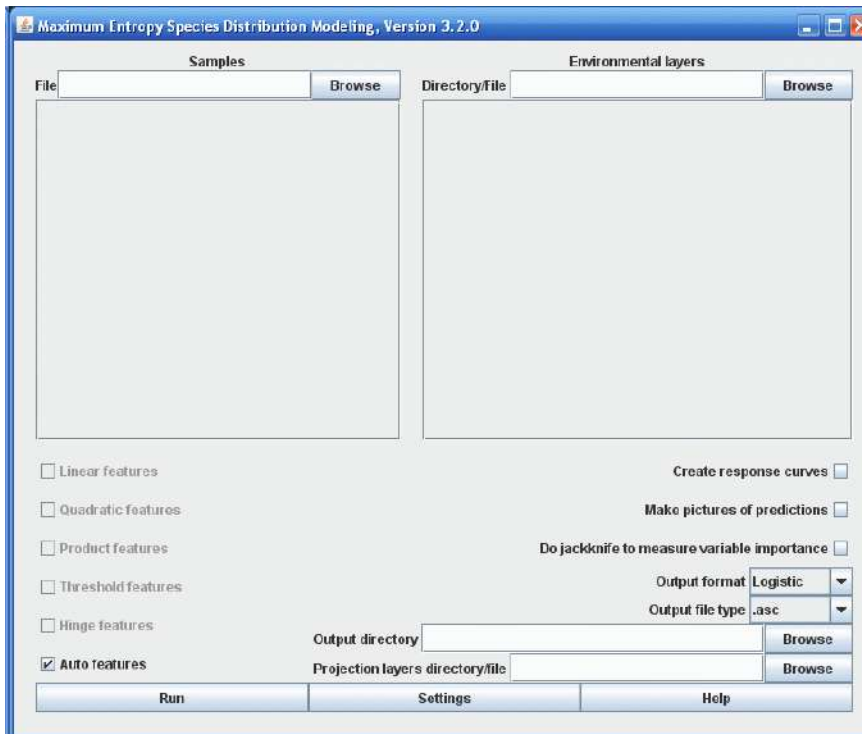
Натисніть «Запустити». Після завершення процесу всі растрові шари будуть обрізані і збережені у ASC-форматі.

На наступному етапі нам потрібно завантажити програмне забезпечення Maxent. Програмне забезпечення містить файл maxent.jar, який може бути запущений на будь-якому

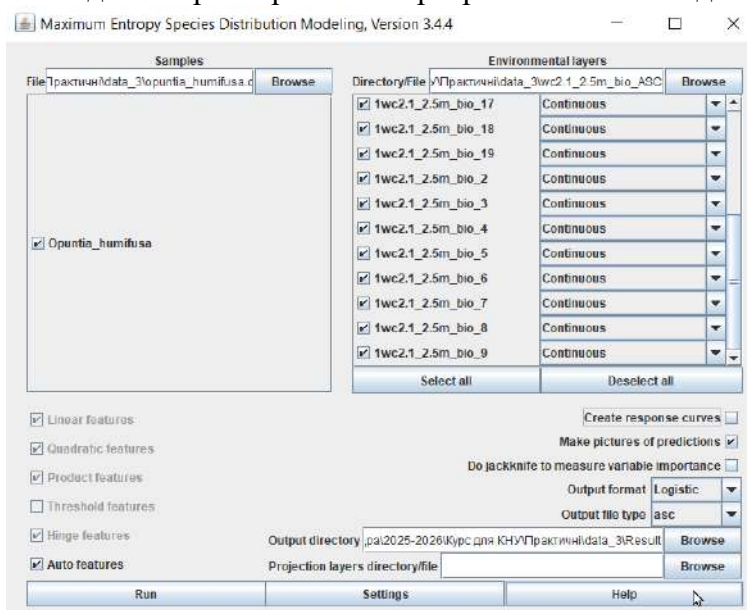
комп'ютері, у якому є версія Java 1.4 або вище. Сам Maxent та документацію можна завантажити за адресою [biodiversityinformatics.amnh.org/open\\_source/maxent](http://biodiversityinformatics.amnh.org/open_source/maxent).

Отримати Java можна за адресою <http://www.oracle.com/technetwork/java/javase/downloads/index.html>

Якщо ви використовуєте Microsoft Windows, просто натисніть на файл maxent.bat. З'явиться наступне вікно:



Щоб запустити процес, потрібно надати файл, що містить точки знахідок біологічного виду (“зразки”), папку, що містить шари з параметрами навколишнього середовища та вихідну папку. У нашому випадку точки знаходяться у файлі “oruntia\_humifusa.csv”, шари параметрів середовища в папці “wc2.1\_2.5m\_bio\_ASC” вихідні результати зберігатимуться в папці “Result”, яку потрібно попередньо створити в робочому каталозі. Коли ви задаєте змінні середовища, пам’ятайте, що потрібна папка, яка містить їх, а не самі файли. Окрім цього оберіть логістичний формат виведення даних. Після введення необхідних параметрів вікно програми стане виглядати наступним чином:



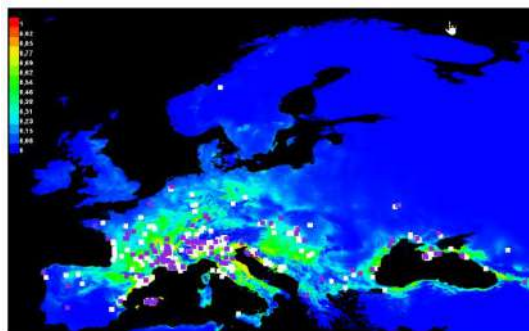
Якщо дані є дублікатами (кілька записів для одного виду в одній комірці), то вони видаляються; це правило можна відключити, натиснувши на кнопку «Налаштування» та відключити «Видалити повторені записи». Також у цій вкладці вкажіть у полі «random test percentage» 25.

Екологічні предиктори у цій роботі є континуальними і описують біокліматичні параметри оточуючого середовища. Усі растри повинні мати однакове географічне охоплення та розмір комірок (заголовки файлів повинні бути строго одноманітними). Якщо використовувати категоріальні змінні, наприклад клас потенційної рослинності, категорії мають бути вказані цифрами, а не буквами або словами. У цьому випадку необхідно вказати програмі, яка зі змінних є категоріальною. Окрім цього оберіть логістичний формат виведення даних.

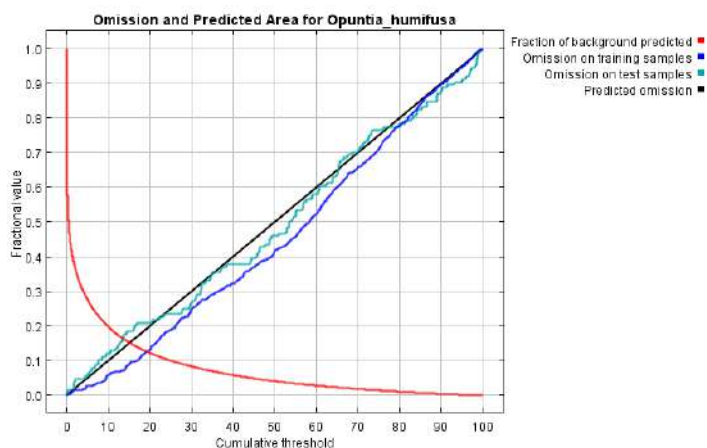
Для запуску моделювання натисніть кнопку "Run". З'явиться індикатор прогресу, що описує поточні кроки роботи програми. Після завершення процесу буде створено кілька вихідних файлів, основним з яких є "opuntia\_humifusa.html". В кінці цього файлу також містяться посилання на інші результати. За замовчуванням вихідні результати у вигляді HTML-сторінки містять графічний результат моделі, застосованої до заданих параметрів навколишнього середовища:

**Pictures of the model**

This is a representation of the Maxent model for *Opuntia\_humifusa*. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

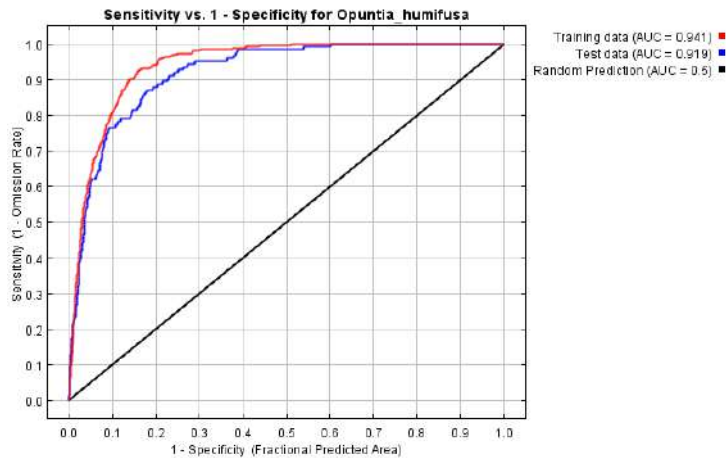


Результуюче зображення кольорами показує розрахункову ймовірність того, що умови для знахідки відповідні: червоним показується висока ймовірність відповідних умов для виду; зеленим - умови, схожі на ті, в яких знаходиться вид; відтінки синього - малоімовірні умови. Цифра "25", яку ми ввели як відсоток випадкових тестових даних ("random test percentage"), говорить програмі, що вона повинна випадковим чином відібрати 25% знахідок і відкласти їх для тестування. Це дозволяє зробити статистичний аналіз точності. Часто для аналізу використовується поріг для бінаризації передбачення, умови вважаються придатними, якщо передбачення вище порогу, і непридатними, якщо нижче. Перший графік показує, як змінюється омісія (помилки моделі) і передбачена територія за тестовими і тренувальними точками в залежності від кумулятивного порогу:



За цим графіком можна сказати, що омісія за тестовими точках досить добре збігається з передбаченою динамікою омісії, розрахованої для тестових даних, отриманих з самого розподілу Maxent. Передбачена омісія є прямою лінією по визначенню кумулятивного формату.

Наступний графік показує receiver operating curve (ROC) для тренувальних і тестових даних.



Червоний і синій графік будуть збігатися, якщо для тренування і тестування використовуються одні й ті ж дані. Якщо дані розділені на дві частини, одна з яких використовується для тренування, а друга для тестування, то червона крива (тренування) показує значення AUC вище, ніж синя (тестування). Червона крива показує, наскільки добре модель описує тренувальні дані ("fit"). Синя лінія показує, наскільки добре модель описує тестові дані і є реальним тестом передбачувальної здатності моделі. Чорна лінія показує ситуацію, яку можна було б очікувати, якби надійність прогнозів моделі була на випадковому рівні. Якщо синя або червона лінії знаходяться нижче чорної, це означає, що рівень достовірних прогнозів моделі навіть нижче, ніж випадковий. Чим ближче до верхнього лівого кута знаходиться синя лінія, тим краще модель передбачає знахідки, що містяться в тестовій вибірці. Оскільки у нас є тільки дані про знахідки, але немає даних про відсутність, замість помилки омісії (частка відсутніх, передбачена як зустрічі) використовується "fractional predicted area" (частка від площі території дослідження, займана зустрічами). Важливо відзначити, що значення AUC мають тенденцію бути вище для видів з невеликими ділянками території дослідження, описуваної шарами даних. Це не обов'язково означає, що модель краща, це всього лише артефакт показника AUC.

Якщо є дані для тестування, програма автоматично підрахує статистичну значущість передбачення, використовуючи біноміальний тест на омісію. Для *Opuntia humifusa* отримаємо таку таблицю:

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.023	Fixed cumulative value 1	0.427	0.008	0.016	2.767E-36
5.000	0.084	Fixed cumulative value 5	0.275	0.027	0.056	1.129E-62
10.000	0.148	Fixed cumulative value 10	0.198	0.059	0.121	0E0
0.332	0.010	Minimum training presence	0.511	0.000	0.016	3.078E-26
16.661	0.239	10 percentile training presence	0.141	0.099	0.202	0E0
19.357	0.276	Equal training sensitivity and specificity	0.126	0.126	0.210	0E0
13.773	0.201	Maximum training sensitivity plus specificity	0.162	0.072	0.177	0E0
13.368	0.195	Equal test sensitivity and specificity	0.165	0.072	0.161	0E0
11.779	0.172	Maximum test sensitivity plus specificity	0.179	0.067	0.129	0E0
3.910	0.069	Balance training omission, predicted area and threshold value	0.302	0.016	0.048	2.516E-56
9.971	0.147	Equate entropy of thresholded and original distributions	0.198	0.059	0.121	0E0

Природне застосування моделювання - спроба відповісти на питання, яка з змінних важливіша для виду? Існує кілька способів дати відповідь на це питання за допомогою Maxent.

В процесі тренування моделі Maxent відстежується, які змінні середовища роблять внесок в побудову моделі. Кожен крок алгоритму Maxent збільшує приріст моделі, змінюючи коефіцієнт для однієї з функцій градієнта середовища; програма призначає збільшення приросту тієї змінної або змінним середовища, від якої залежить ця функція. В кінці процесу тренування відбувається перерахунок приростів в відсотки і виходить середня колонка в таблиці:

## Analysis of variable contributions

The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is reevaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution	Permutation importance
lwc2.1_2.5m_bio_11	50.3	62
lwc2.1_2.5m_bio_7	13.4	6.7
lwc2.1_2.5m_bio_18	13.2	11.8
lwc2.1_2.5m_bio_1	6	0.2
lwc2.1_2.5m_bio_19	3	1.3
lwc2.1_2.5m_bio_2	2.8	1.3
lwc2.1_2.5m_bio_6	2.4	0.8
lwc2.1_2.5m_bio_3	1.3	2.4
lwc2.1_2.5m_bio_17	1.3	0.3
lwc2.1_2.5m_bio_14	1.2	2
lwc2.1_2.5m_bio_13	1.1	1.2
lwc2.1_2.5m_bio_16	1	2.6
lwc2.1_2.5m_bio_8	0.6	0.5
lwc2.1_2.5m_bio_10	0.6	2.5
lwc2.1_2.5m_bio_15	0.5	0.4
lwc2.1_2.5m_bio_9	0.5	1.8
lwc2.1_2.5m_bio_4	0.3	1.8
lwc2.1_2.5m_bio_5	0.2	0.4
lwc2.1_2.5m_bio_12	0	0

Ці процентні внески визначаються тільки евристично і залежать від конкретного шляху, по якому пішов код Maxent, щоб досягти оптимального рішення. Інший алгоритм міг би прийти до того ж рішення, використовуючи інший шлях, який би закінчився іншими відсотками вкладу.

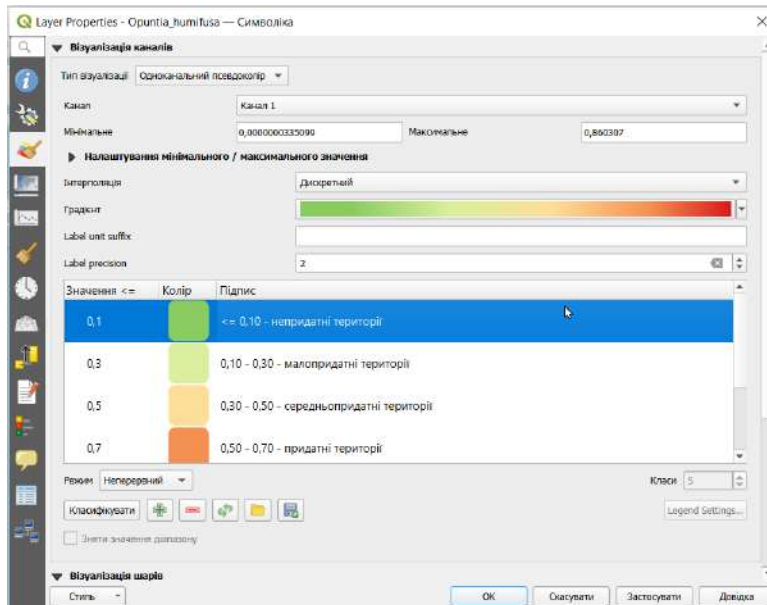
Права колонка в таблиці показує другий показник вкладу змінної - важливість при пермутації (permutation importance). Цей показник залежить тільки від остаточної моделі, а не шляху, який був пройдений, щоб її отримати. Внесок кожної змінної визначається за допомогою випадкової зміни значень цієї змінної у тренувальних точках (і знахідок, і фону), а також вимірювання зменшення тренувальної AUC. Значне зменшення свідчить про те, що модель сильно залежить від цієї змінної. Значення нормалізовані і показуються в процентах.

Отже, моделювання та інтерпретацію його результатів завершено. Можемо підготувати карту за підсумками моделювання.

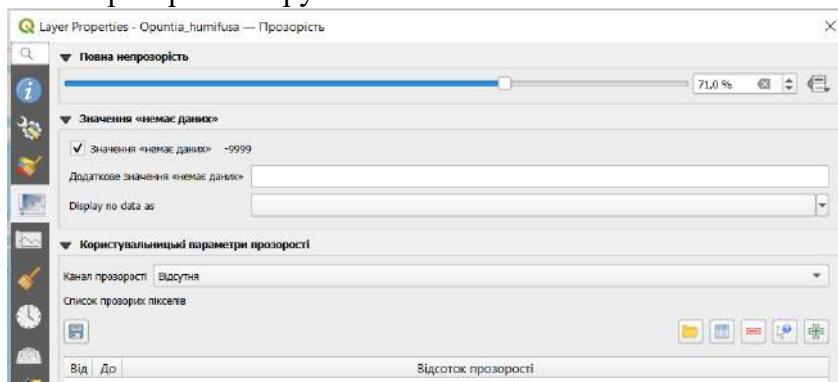
Додайте в проект Q-GIS точковий шар знахідок, растровий шар результатів *Oruntia humifusa.asc* та картооснову. Для растрового шару задайте систему координат проекту WGS 84 і збережіть проект. Для більш адекватного відображення даних змініть проекцію проекту на EPSG: 3857.



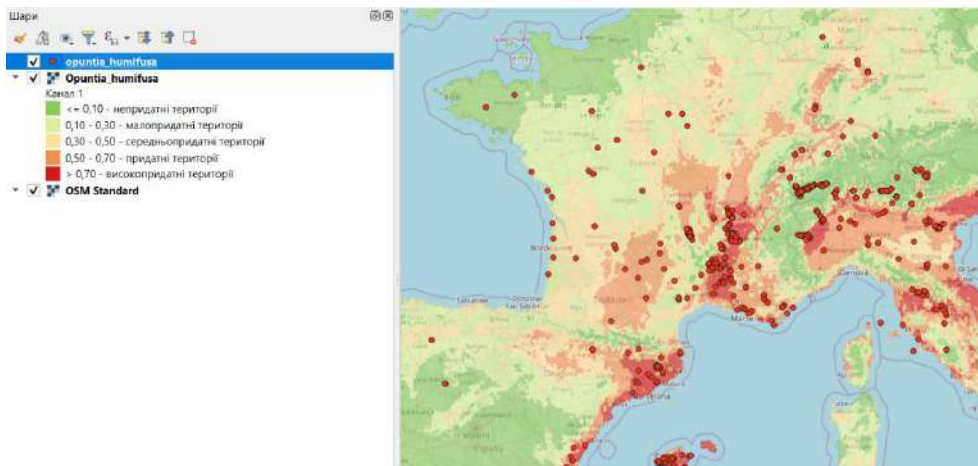
Як бачимо, значення комірок растру коливаються від 0 (непридатні для виду території) до 0,86 (високопридатні території). Ми можемо візуалізувати ці дані за допомогою псевдоколірів. Оберіть «Одноканальний псевдоколір» та дискретний алгоритм інтерполяції. Для створення зон розподілу потенційного ареалу прийнято використовувати наступні порогові значення: < 0,1 – непридатні території, 0,1 – 0,3 – малопридатні, 0,3–0,50 – середньопридатні, 0,5 – 0,7 – придатні, 0,7 – 0,9 – високопридатні та > 0,9 – найкращі.



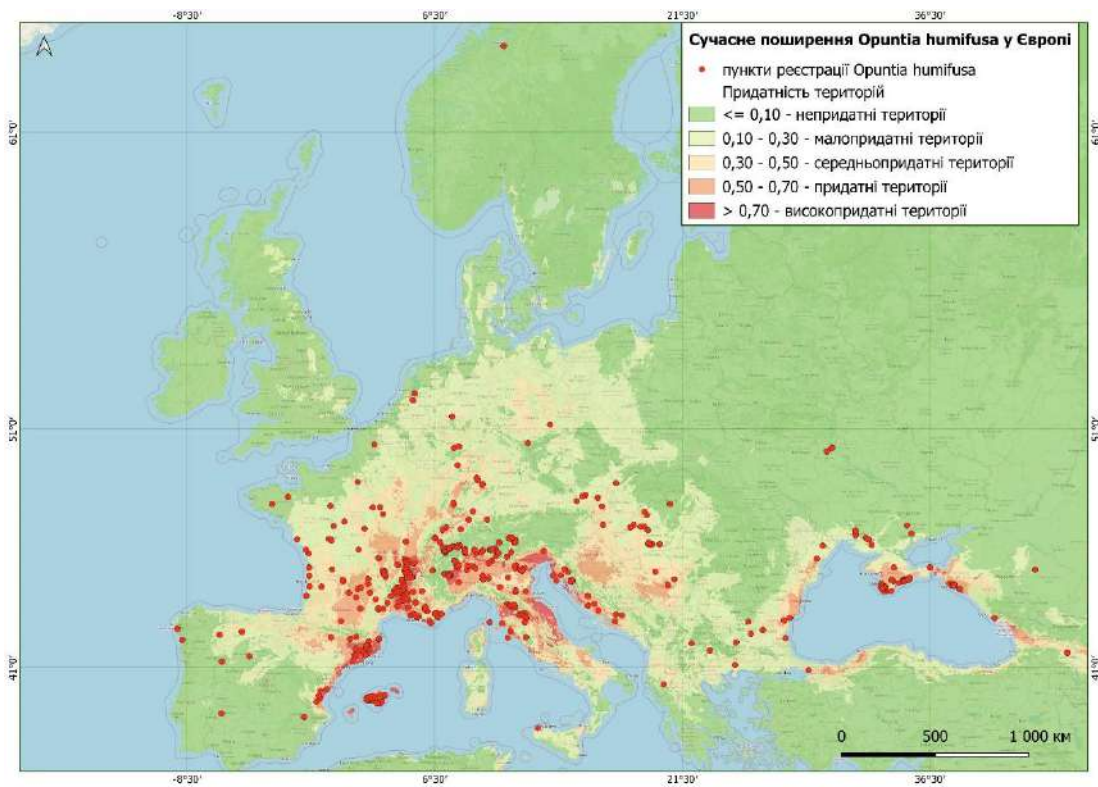
Налаштуйте також прозорість шару:



Також можете розмістити точковий шар над растровим, щоб показати на карті використані локалітети.



Дані готові до налаштування макету карти. Створіть новий макет карти та підготуйте його до друку.



**Завдання для самостійної роботи.** Оберіть на ваш розсуд один із видів рослин або тварин, занесених до Червоної книги України. Завантажте з GBIF дані щодо його поширення у Європі. Майте на увазі, що широкорозповсюджені масові і добре досліджені види можуть мати десятки тисяч записів у GBIF. Надто великі масиви даних ви не зможете обробити в ексель. У такому випадку можна зменшити територію інтересу в процесі завантаження даних. Наприклад, обмежитись лише Україною.

Завантажте екологічні предиктори з WorldClim. Окрім біокліматичних змінних є можливість завантажити і використовувати й інші параметри середовища: мінімальна та максимальна температура, середня температура, кількість опадів, сонячна радіація, швидкість вітру, евопотранспірація.

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
minimum temperature (°C)	<a href="#">tmin 10m</a>	<a href="#">tmin 5m</a>	<a href="#">tmin 2.5m</a>	<a href="#">tmin 30s</a>
maximum temperature (°C)	<a href="#">tmax 10m</a>	<a href="#">tmax 5m</a>	<a href="#">tmax 2.5m</a>	<a href="#">tmax 30s</a>
average temperature (°C)	<a href="#">tavg 10m</a>	<a href="#">tavg 5m</a>	<a href="#">tavg 2.5m</a>	<a href="#">tavg 30s</a>
precipitation (mm)	<a href="#">prec 10m</a>	<a href="#">prec 5m</a>	<a href="#">prec 2.5m</a>	<a href="#">prec 30s</a>
solar radiation (kJ m <sup>-2</sup> day <sup>-1</sup> )	<a href="#">srad 10m</a>	<a href="#">srad 5m</a>	<a href="#">srad 2.5m</a>	<a href="#">srad 30s</a>
wind speed (m s <sup>-1</sup> )	<a href="#">wind 10m</a>	<a href="#">wind 5m</a>	<a href="#">wind 2.5m</a>	<a href="#">wind 30s</a>
water vapor pressure (kPa)	<a href="#">vapr 10m</a>	<a href="#">vapr 5m</a>	<a href="#">vapr 2.5m</a>	<a href="#">vapr 30s</a>

[Data management platform](#)

Також доступний для завантаження шар висот над рівнем моря.

variable	10 minutes	5 minutes	2.5 minutes	30 seconds
Elevation	<a href="#">elev 10m</a>	<a href="#">elev 5m</a>	<a href="#">elev 2.5m</a>	<a href="#">elev 30s</a>

Також ви можете використати шар цифрової моделі рельєфу, наприклад SRTM. З нього, в свою чергу, можна розрахувати ухил та експозицію схилів. Практично безмежним джерелом даних є

космічні знімки та продукти створені на їх основі: типи земельного покриття, вегетаційні індекси, тощо. Всі ці дані можна використовувати для потреб моделювання.

Підготуйте завантажені дані (як описано вище) і проведіть моделювання поширення виду.

**Підготуйте звіт за результатами моделювання, який має включати карту поширення виду, статистичні оцінки моделі та їх інтерпретацію.**

#### **Рекомендована література**

1. Титар В.М. Аналіз ареалів видів: підхід, заснований на моделюванні екологічної ніші. Вістник зоології. Окремий випуск. 2011. №25. 96 с.
2. Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 2006. №190, P. 231–259.
3. Neftalí Sillero & A. Márcia Barbosa (2021) Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, 2021. №35:2. P. 213-226.
4. Spatial modelling for koalas in South East Queensland: Report version 4.0. Koala Habitat Areas (KHA) v4.0, Locally Refined Koala Habitat Areas (LRKHA) v4.0, Koala Priority Areas (KPA) v1.0, Koala Habitat Restoration Areas (KHRA) v1.0. Brisbane: Department of Environment and Science, Queensland Government, 2023. 107 p.
5. Araújo, M.B., et al., 2019. Standards for distribution models in biodiversity assessments. *Science Advances*, 5, eaat4858
6. Feng, X., et al., 2019. A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution*, 3, 1382–1395.
7. Sofaer, H.R., et al., 2019. Development and delivery of species distribution models to inform decision-making. *BioScience*, 69 (7), 544–557.