

Міністерство освіти та науки України
Національний університет водного господарства та
природокористування
Кафедра екології, технології захисту навколишнього середовища та
лісового господарства

05-02-492М

МЕТОДИЧНІ ВКАЗІВКИ

до виконання практичних робіт з навчальної дисципліни «Статистичні
методи обробки даних»
для здобувачів вищої освіти третього (освітньо-наукового) рівня
за освітньою програмою «Екологія» спеціальності 101 «Екологія»
денної та заочної форм навчання

Рекомендовано
науково-методичною радою з якості
НП агроекології та землеустрою
Протокол № 11 від 28.01.2025 р.

Рівне – 2025

Методичні вказівки до виконання практичних робіт з навчальної дисципліни «Статистичні методи обробки даних» для здобувачів вищої освіти третього (освітньо-наукового) рівня за освітньою програмою «Екологія» спеціальності 101 «Екологія» денної та заочної форм навчання [Електронне видання] / Клименко М. О., Буднік З. М. – Рівне : НУВГП, 2025. – 39 с.

Укладачі: Клименко М. О., доктор сільськогосподарських наук, професор, завідувач кафедри екології, технології захисту навколишнього середовища та лісового господарства;
Буднік З. М. – кандидат сільськогосподарських наук, доцент кафедри екології, технології захисту навколишнього середовища та лісового господарства.

Відповідальний за випуск: Клименко М. О., доктор сільськогосподарських наук, професор, завідувач кафедри екології, технології захисту навколишнього середовища та лісового господарства.

Керівник групи забезпечення освітньо-професійної програми третього (освітньо-наукового) рівня за освітньою програмою «Екологія» спеціальності 101 «Екологія» д.біо.н., професор Бедункова О. О.

© З. М. Буднік, 2025
© Національний університет
водного господарства та
природокористування, 2025

Зміст

Передмова	3
Практична робота №1. Ознайомлення з інструментами аналізу даних	5
Практична робота №2. Обчислення описових статистик	11
практична робота №3:Проведення t-тесту та аналізу χ^2	18
практична робота №4. Нелінійний кореляційний аналіз.....	22
практична робота №5. Лінійний регресійний аналіз.....	26
практична робота № 6. Двохфакторний дисперсійний аналіз із даними без повторень.....	31
практична робота №7. Аналіз часових рядів та прогнозування екологічних змін	35
Список використаних літературних джерел	38

ПЕРЕДМОВА

Практичні роботи є важливою складовою навчального процесу в рамках дисципліни «Статистичні методи обробки даних». Їх мета – забезпечити здобувачів вищої освіти спеціальності 101 «Екологія» знаннями та навичками, необхідними для ефективного використання сучасних статистичних методів і програмного забезпечення у дослідженні екологічних систем та аналізі великих обсягів даних.

Екологія як наука часто оперує даними, що відображають складні природні процеси та взаємодії. Застосування статистичних методів дозволяє систематизувати ці дані, виявити закономірності, оцінити вплив різних факторів на стан довкілля, а також будувати прогнози щодо змін у природних екосистемах.

Дані методичні вказівки створені для спрямування аспірантів у виконанні практичних завдань. Вони містять:

- Мету та завдання кожної практичної роботи.
- Вихідні дані для виконання завдань.
- Опис методів та інструментів, що використовуються.
- Покрокові інструкції для виконання кожного етапу роботи.
- Контрольні запитання для перевірки розуміння матеріалу.

Практичні роботи орієнтовані на освоєння студентами таких ключових аспектів:

1. Збір, організація та обробка екологічних даних.

2. Застосування описових статистик для оцінки стану екологічних систем.
3. Побудова та перевірка статистичних гіпотез.
4. Використання багатовимірного аналізу для моделювання складних екологічних процесів.
5. Аналіз просторових та часових змін у природних системах.

Методичні вказівки враховують різний рівень підготовки аспірантів, тому практичні завдання побудовані за принципом поступового ускладнення: від ознайомлення з базовими інструментами до застосування складних статистичних методів у реальних екологічних дослідженнях.

Виконання практичних робіт дозволить аспірантам:

- Освоїти роботу з програмним забезпеченням (R, Python, SPSS тощо).
- Розвинути аналітичні навички та вміння працювати з великими масивами екологічної інформації.
- Отримати практичний досвід, необхідний для проведення наукових досліджень у сфері екології.

При виконанні практичних робіт здобувачами вищої освіти набуваються загальні та фахові компетентності:

ЗК 02. Здатність розв'язувати комплексні проблеми на основі системного наукового та загальнокультурного світогляду із дотриманням принципів професійної етики та академічної доброчесності.

СК01. Здатність виконувати оригінальні дослідження, досягати наукових результатів, які створюють нові знання у сфері екології та дотичних до неї міждисциплінарних напрямках, оцінювати та забезпечувати якість виконуваних досліджень.

СК03. Здатність застосовувати сучасні інструменти, електронні інформаційні ресурси, спеціалізоване програмне забезпечення у науковій та навчальній діяльності, зокрема для моделювання процесів та прийняття оптимальних рішень у сфері екології, охорони природи та раціонального природокористування.

Виконання практичних робіт забезпечує набуття таких програмних результатів навчання:

ПР01. Глибоко розуміти концептуальні принципи та методологію природничих наук, формулювати і перевіряти гіпотези, використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних

досліджень і математичного та/або комп'ютерного моделювання з метою розв'язання значущих наукових та науково-прикладних проблем екології.

ПРО6. Застосовувати сучасні інструменти та технології пошуку оброблення й аналізу інформації з проблем екології та дотичних питань, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи

Ці методичні вказівки є орієнтиром для самостійної та командної роботи аспірантів, сприяючи їхньому професійному зростанню та формуванню компетентностей, необхідних для вирішення сучасних екологічних проблем.

ПРАКТИЧНА РОБОТА №1

ОЗНАЙОМЛЕННЯ З ІНСТРУМЕНТАМИ АНАЛІЗУ ДАНИХ

Мета роботи: ознайомитись із інструментами аналізу даних. Розвинути початкові навички роботи з такими інструментами, як R, Python (бібліотеки NumPy, Pandas, Matplotlib, SciPy), SPSS або Statistica.

Основні поняття

Статистичний аналіз даних – важлива складова для будь-якої діяльності, де є великі масиви інформації, розрахунки та моделювання. Статистичні програми незамінні для розробників, експериментаторів, наукових співробітників, студентів вищих навчальних закладів, викладачів.

Система екологічної інформації заснована на створенні комп'ютеризованої мережі збору, обробки, зберігання і обміну екологічної інформації. Вона є основою існування системи моніторингу і базується на виробленні спеціальних програм; збиранні, обробці первинної інформації від станцій спостереження та інших джерел, обміні з іншими інформаційними мережами, передачі інформації та ін.; моделюванні процесів екологічного впливу; створенні екологічних карт; прогнозуванні майбутніх наслідків впливу на навколишнє середовище.

Найбільш поширені програмні продукти для статистичної обробки даних:

1. Python

Звичайно, куди ж без Python. Сьогодні він посідає одне з перших місць за популярністю серед мов програмування, Python обов'язковий інструмент для аналітиків даних. Для тих, хто не знає, — це мова програмування з відкритим вихідним кодом, її вважають максимально універсальною, простою у використанні, саме тому її часто використовують новачки в програмуванні. Але також це незамінний інструмент для дата-аналітиків.

Python може реалізовувати будь-які завдання у сфері дата-аналітики завдяки багатій екосистемі бібліотек: від попереднього опрацювання даних, візуалізації та статистичного аналізу до розгортання моделей машинного та глибокого навчання.

2. Мова програмування R

Ще одна мова програмування, яка необхідна дата-аналітику. Важливо зазначити, що вона не настільки популярна, як Python, але водночас дуже важлива для роботи з даними.

R — це мова з відкритим вихідним кодом, яка була спеціально створена для науки про дані та передових обчислень. Вона дуже популярна у сфері фінансів та академічних досліджень. Сьогодні R — ідеальна мова для маніпулювання даними, обробки, візуалізації, статистичних обчислень і машинного навчання. Ця мова може похвалитися великим ком'юніті та широкою колекцією спеціалізованих бібліотек для аналізу даних і машинного навчання.

Якщо порівнювати її з Python, то новачкам вона може здатися набагато складнішою, проте, можливо, і не відразу, але вам варто її вивчити.

3. SQL

Далі в нас іде ще одна мова програмування — SQL. Це предметно-орієнтована мова, за допомогою якої програмісти можуть працювати з базами даних, а також редагувати і витягувати з них необхідні дані. Для того щоб стати аналітиком даних, без ґрунтовних знань баз даних і SQL точно не обійтись. Це дасть вам змогу працювати з різними реляційними базами даних, включно з такими популярними системами, як SQLite, MySQL і PostgreSQL. Завдяки простому синтаксису, SQL дуже зручний у вивченні порівняно з іншими мовами.

4. Excel

Microsoft Excel — це класика, яка знайома нам ще зі школи. І хоча більшість фахівців ставляться до цього інструменту доволі скептично,

він все ніяк не може покинути список сильних інструментів для дата-аналітиків. Так, сьогодні вже є безліч більш сучасних, просунутих і зручних програм, але Excel, як і раніше, залишається інструментом для багатьох повсякденних завдань аналізу даних.

Excel поєднує в собі потужні можливості зі зручним інтерфейсом, що робить його непоганим варіантом для досвідчених аналітиків даних і не супер технічно підкованих користувачів. Це також чудовий інструмент завдяки його плавній інтеграції зі стеком Microsoft BI, включно з PowerBI.

5. PowerBI

Power BI — це хмарне рішення для бізнес-аналітики, яке дає змогу об'єднувати різні джерела даних, аналізувати їх і представляти аналіз даних за допомогою візуалізацій, звітів і панелей моніторингу.

Серед основних функцій Power BI:

- Пряме підключення до Excel і проста інтеграція з іншими продуктами Microsoft.
- Можливість стискати дані та витягувати цінну інформацію з великих наборів даних.
- Можливість налаштування з використанням R і Python.
- Використання Power Query для імпорту, фільтрації та перекладу даних.
- Адаптивний і зручний користувацький інтерфейс.

Кожен аналітик даних-початківець повинен вільно володіти хоча б одним інструментом BI. Power BI, без сумніву, один із найкращих кандидатів.

6. Tableau

Tableau — ще один популярний інструмент візуалізації даних, який дає змогу одночасно аналізувати дані з кількох джерел.

Tableau вважають одним із найкращих інструментів BI. Він використовується провідними компаніями для вилучення інформації з величезних обсягів необроблених даних. Завдяки інтуїтивно зрозумілій і потужній платформі як технічні, так і нетехнічні користувачі можуть створювати візуальні панелі моніторингу та робочі листи.

У багатьох аспектах Tableau пропонує ті самі можливості, що й Power BI. Однак він вважається складнішим інструментом для новачків. Тому, якщо вас цікавить лише побудова простих діаграм, вам слід обирати менш складні варіанти, але надалі, коли ви наберетеся досвіду, вам все одно доведеться вивчити цю програму.

7. Google BigQuery

І завершує наш список Google BigQuery. Це хмарна платформа для аналітики великих даних, що належить до екосистеми Google Cloud. Вона дає змогу користувачам зберігати й аналізувати великі обсяги даних у реальному часі.

BigQuery особливо цінний для компаній, які проводять великомасштабний аналіз даних, особливо тих, які вже використовують сервіси Google Cloud. Крім аналізу великих даних цей інструмент часто використовують для бізнес-аналітики, аналізу поведінки користувачів, прогнозування продажів і аналізу соціальних медіа. Серед його головних особливостей — масштабованість, швидкість, підтримка SQL, інтеграція з іншими сервісами Google Cloud.

Завдання роботи

1. Ознайомитися з основними можливостями обраного програмного забезпечення.
2. Налаштувати робоче середовище для аналізу даних.
3. Виконати базові операції з імпорту, перегляду та первинної обробки екологічних даних.
4. Виконати візуалізацію базових статистичних показників.

Хід виконання роботи

1. Вступна частина
 - Ознайомлення з метою роботи та переліком доступних програмних засобів.
 - Огляд інтерфейсу вибраного інструменту (наприклад, RStudio для R, Jupyter Notebook для Python).
 - Обговорення переваг та особливостей використання інструменту для екологічного аналізу.
2. Підготовка до роботи
 - Завантаження програмного забезпечення (за необхідності):
 - **R**: Встановлення R і RStudio.
 - **Python**: Встановлення Python і бібліотек NumPy, Pandas, Matplotlib.
 - **SPSS** або **Statistica**: Ознайомлення з доступними функціями та налаштування робочого середовища.
3. Імпорт даних
 - Підготовка екологічного датасету (наприклад, дані про концентрацію забруднювачів у воді або ґрунті).

- Виконання імпорту даних:
 - R: `read.csv()` для завантаження файлу CSV.
 - Python: Використання `pandas.read_csv()`.
 - SPSS/Statistica: Імпорт файлу через інтерфейс.

4. Аналіз структури даних

- Перегляд основної інформації про датасет:
 - R: `str()` та `summary()`.
 - Python: `df.info()` та `df.describe()`.
 - SPSS: Таблиця частот і основні статистичні показники.

5. Візуалізація даних

- Побудова простих графіків для аналізу даних:
 - R: `hist()`, `boxplot()`.
 - Python: `matplotlib.pyplot.hist()` і `seaborn.boxplot()`.
 - SPSS: Побудова гістограм і діаграм через меню "Графіки".

6. Збереження результатів

- Збереження модифікованого датасету після первинної обробки.
 - R: `write.csv()`.
 - Python: `df.to_csv()`.
 - SPSS: Збереження через опцію "Сохранить как".

Опис вихідних даних

Для виконання завдань практичної роботи студентам надається екологічний датасет, який містить інформацію про стан довкілля в певному регіоні. Датасет розроблений для демонстрації базових статистичних операцій та візуалізації.

Формат даних:

- Тип файлу: CSV (Comma-Separated Values).
- Назва файлу: `environmental_data.csv`.
- Джерело даних: реальні або симульовані дані екологічного моніторингу.

Структура датасету

Колонка	Тип даних	Опис
Location	Текстовий	Назва місця збору даних (наприклад, "Станція 1", "Станція 2").
Date	Дата	Дата проведення вимірювання (формат: YYYY-MM-DD).
Temperature	Числовий	Температура повітря, °C.

pH_Water	Числовий	Рівень кислотності води (pH).
Oxygen_Concentration	Числовий	Концентрація кисню у воді, мг/л.
Pollutant_Level	Числовий	Рівень забруднення повітря, мкг/м ³ .
Vegetation_Cover	Числовий	Відсоток рослинного покриву (%) у зоні дослідження.

Приклад фрагмента датасету

Location	Date	Temperature	pH_Water	Oxygen_Concentration	Pollutant_Level	Vegetation_Cover
Станція 1	2024-01-15	12.5	7.2	8.5	35	70
Станція 2	2024-01-15	14.0	6.8	7.8	50	60
Станція 3	2024-01-15	13.2	7.0	8.2	40	65

Пояснення до колонок

1. **Location** – Інформація про точку збору даних, важлива для географічного аналізу.
2. **Date** – Дата збору даних, необхідна для аналізу часових тенденцій.
3. **Temperature** – Екологічний параметр, який впливає на фізико-хімічні процеси в екосистемі.
4. **pH_Water** – Ключовий показник для оцінки стану водного середовища.
5. **Oxygen_Concentration** – Показник стану водної екосистеми (розчинений кисень).
6. **Pollutant_Level** – Показує ступінь забруднення повітря.
7. **Vegetation_Cover** – Дає уявлення про біологічну продуктивність регіону.

Додаткові матеріали для аналізу

- **Мета використання:** Датасет дозволяє виконати базовий описовий аналіз, провести візуалізацію даних і зрозуміти структуру екологічної інформації.
- **Рекомендації:**

○ Використовуйте частину датасету для навчання, щоб освоїти методи імпорту та обробки даних.

○ Для роботи з реальними даними зверніть увагу на якість (наявність пропусків, аномалії тощо).

З цими вихідними даними студенти зможуть навчитися основам роботи з інструментами для аналізу екологічних даних.

Звіт про виконання роботи

Звіт повинен містити:

1. Мету та завдання роботи.
2. Опис датасету (структура, кількість змінних та спостережень).
3. Скріншоти або результати виконання завдань (графіки, виведення статистичних функцій).
4. Відповіді на контрольні запитання.

Контрольні запитання

1. Які основні функції обраного інструменту для аналізу даних?
2. Як здійснюється імпорт даних у вибраному програмному забезпеченні?
3. Що таке описова статистика, і як її обчислити у вашому інструменті?
4. Які типи графіків можна побудувати для первинного аналізу даних?

ПРАКТИЧНА РОБОТА №2 ***ОБЧИСЛЕННЯ ОПИСОВИХ СТАТИСТИК***

Мета роботи: ознайомити студентів із методами обчислення описових статистик для аналізу екологічних даних. Розвинути навички інтерпретації статистичних показників, що характеризують розподіл, варіативність та взаємозв'язки між екологічними параметрами.

Основні поняття

До первинних описових статистик або дескриптивних статистик зазвичай відносять числові характеристики розподілу ознаки, що вимірюється на вибірці. Кожна така характеристика відображає в одному числовому значенні властивість розподілу безлічі результатів вимірювання: з точки зору їх розташування на числовій осі або з точки зору їх мінливості. Основне призначення кожної з первинних описових статистик - заміна безлічі значень ознаки, вимірюваної на вибірці, одним числом. Компактний опис групи за допомогою описових статистик

дозволяє інтерпретувати результати вимірювань, зокрема, шляхом порівняння результатів різних груп.

Описові статистики використовують три основні методи агрегування даних:

1. Представлення у вигляді таблиці
2. Графічне зображення
3. Розрахунок статистичних показників

Розрахунок статистичних показників можна умовно розділити на: міри центральної тенденції, міри розкиду, квантилі і показники форми розподілу.

Міри центральної тенденції

Міра центральної тенденції - це число, що характеризує вибірку за рівнем вираженості ознаки, що вимірюється.

Існує три способи визначення "центральної тенденції", кожному з яких відповідає своя міра:

- Мода
- Медіана
- Вибіркове середнє

Міри розкиду

Міри розкиду (мінливості) застосовуються в психології для чисельного вираження величини міжіндивідуальної варіації ознаки і показують, наскільки добре дані значення представляють дану сукупність.

В якості найбільш використовуваних заходів мінливості слід назвати:

- Мінімальне
- Максимальне
- Розмах
- Міжквартильний розмах
- Дисперсія
- Стандартне відхилення
- Коефіцієнт варіації

Показники форми розподілу

Більшість розрахунків кореляції вимагають нормальності розподілу результатів. Для того що б визначити "нормальність" існують показати:

- Асиметрія
- Ексцес

Завдання роботи

1. Виконати описовий аналіз заданого екологічного датасету.
2. Обчислити базові статистичні показники: середнє значення, моду, медіану, мінімум, максимум, дисперсію, стандартне відхилення, коефіцієнт варіації.
3. Провести графічну візуалізацію результатів (гістограма, коробчатий графік).
4. Інтерпретувати отримані результати.

Вихідні дані

Датасет із даними екологічного моніторингу.

Назва датасету: *environmental_data.csv*.

Колонки:

- *Location*: назва станції збору даних.
- *Temperature*: температура повітря, °C.
- *pH_Water*: рівень кислотності води.
- *Oxygen_Concentration*: концентрація кисню, мг/л.
- *Pollutant_Level*: рівень забруднення повітря, мкг/м³.
- *Vegetation_Cover*: відсоток рослинного покриву.

Хід виконання роботи

1. Підготовка до роботи
 - Завантажте програмне забезпечення для аналізу даних (R, Python, SPSS або Excel).
 - Імпортуйте датасет:
 - R: `data <- read.csv("environmental_data.csv")`.
 - Python: `import pandas as pd; data = pd.read_csv("environmental_data.csv")`.
2. Обчислення базових статистичних показників
 1. **Опис кожної змінної (включаючи кількісні показники):**
 - **Середнє значення:**
 - R: `mean(data$Temperature)`
 - Python: `data["Temperature"].mean()`
 - **Медіана:**
 - R: `median(data$Temperature)`
 - Python: `data["Temperature"].median()`
 - **Мода:**
 - R: `table(data$Temperature) |> which.max()`
 - Python: `data["Temperature"].mode()`
 2. **Аналіз варіативності:**

- **Дисперсія:**
 - R: `var(data$Temperature)`
 - Python: `data["Temperature"].var()`
- **Стандартне відхилення:**
 - R: `sd(data$Temperature)`
 - Python: `data["Temperature"].std()`
- **Коефіцієнт варіації:**
 - R: `(sd(data$Temperature) / mean(data$Temperature)) * 100`
 - Python: `(data["Temperature"].std() /`

`data["Temperature"].mean()) * 100`

3. Границі розподілу:

- **Мінімум:**
 - R: `min(data$Temperature)`
 - Python: `data["Temperature"].min()`
- **Максимум:**
 - R: `max(data$Temperature)`
 - Python: `data["Temperature"].max()`

3. Візуалізація даних

1. Побудова гістограми:

- R: `hist(data$Temperature, main="Гістограма температури", xlab="Температура, °C")`
- Python:

```
python
import matplotlib.pyplot as plt
data["Temperature"].hist(bins=10)
plt.title("Гістограма температури")
plt.xlabel("Температура, °C")
plt.ylabel("Частота")
plt.show()
```

2. Коробчатий графік (Boxplot):

- R: `boxplot(data$Temperature, main="Boxplot температури", ylab="Температура, °C")`
- Python:

```
python
import seaborn as sns
sns.boxplot(data[["Temperature"]])
plt.title("Boxplot температури")
plt.ylabel("Температура, °C")
plt.show()
```

4. Інтерпретація результатів

- Проаналізуйте середнє значення, моду та медіану: чи є розподіл симетричним?
- Оцініть варіативність за стандартним відхиленням і коефіцієнтом варіації.
- Інтерпретуйте результати візуалізації:
 - Гістограма дозволяє оцінити розподіл температури (нормальний чи інший).
 - Boxplot допомагає виявити аномальні значення (викиди).

Звіт про виконання роботи

1. Мета роботи.
2. Таблиця з обчисленими статистичними показниками для всіх змінних.
3. Побудовані графіки (гістограми, Boxplot).
4. Відповіді на контрольні запитання.
5. Висновки: короткий аналіз структури та варіативності даних.

Вихідні дані

Для виконання завдання використовується симульований екологічний датасет, що містить показники, зібрані на трьох станціях моніторингу. Дані представлені у вигляді таблиці.

Location	Temperature (°C)	pH_Water	Oxygen_Concentration (mg/L)	Pollutant_Level (µg/m³)	Vegetation_Cover (%)
Станція 1	12.5	7.2	8.5	35	70
Станція 2	14.0	6.8	7.8	50	60
Станція 3	13.2	7.0	8.2	40	65
Станція 4	15.1	7.3	8.0	45	68
Станція 5	11.8	7.1	8.6	30	75

Приклад розрахунків

1. Середнє значення

Середнє значення розраховується як сума всіх значень змінної, поділена на їхню кількість.

Для температури:

$$\text{Середнє значення} = \frac{12.5 + 14.0 + 13.2 + 15.1 + 11.8}{5} = \frac{66.6}{5} = 13.32 \text{ } ^\circ\text{C}$$

2. Медіана

Медіана – це центральне значення упорядкованого ряду.

Для температури (упорядкування): 11.8, 12.5, 13.2, 14.0, 15.1

Медіана = **13.2** °C (середнє значення в ряду).

3. Мода

Мода – це найпоширеніше значення.

У цьому прикладі для температури немає повторюваних значень, тому мода відсутня.

4. Мінімум і максимум

• Мінімум: найменше значення в наборі.

○ Температура: **11.8** °C

• Максимум: найбільше значення в наборі.

○ Температура: **15.1** °C

5. Дисперсія

Дисперсія показує варіативність даних навколо середнього.

Формула:

$$\text{Дисперсія} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Де x_i – кожне значення, \bar{x} – середнє, n – кількість спостережень.

Для температури:

$$\bar{x} = 13.32 \text{ } ^\circ\text{C}$$

$$\begin{aligned} \text{Дисперсія} &= \frac{(12.5 - 13.32)^2 + (14.0 - 13.32)^2 + (13.2 - 13.32)^2 + (15.1 - 13.32)^2 + (11.8 - 13.32)^2}{4} \\ &= \frac{(-0.82)^2 + (0.68)^2 + (-0.12)^2 + (1.78)^2 + (-1.52)^2}{4} \\ &= \frac{0.6724 + 0.4624 + 0.0144 + 3.1684 + 2.3104}{4} \\ &= \frac{6.627}{4} = 1.65675 \text{ } ^\circ\text{C}^2 \end{aligned}$$

6. Стандартне відхилення

Стандартне відхилення – це квадратний корінь із дисперсії:

$$SD = \sqrt{\text{Дисперсія}} = \sqrt{1.65675} \approx 1.29 \text{ } ^\circ\text{C}$$

7. Коефіцієнт варіації

Коефіцієнт варіації показує відносну варіативність (у відсотках):

$$CV = \frac{SD}{\bar{x}} \times 100$$

$$CV = \frac{1.29}{13.32} \times 100 \approx 9.68 \%$$

8. Гістограма температури

Дані діляться на інтервали (наприклад, 11–12, 12–13, ...), будувється частотний розподіл та графік. У Python:

```
python Копіювати Редагувати

import matplotlib.pyplot as plt
data = [12.5, 14.0, 13.2, 15.1, 11.8]
plt.hist(data, bins=5, color="skyblue", edgecolor="black")
plt.title("Гістограма температури")
plt.xlabel("Температура (°C)")
plt.ylabel("Частота")
plt.show()
```

Результати розрахунків для температури

Показник	Значення
Середнє значення	13,32 °C
Медіана	13,2 °C
Мода	Відсутня
Мінімум	11,8 °C
Максимум	15,1 °C
Дисперсія	1,657 °C ²
Стандартне відхилення	1,29 °C
Коефіцієнт варіації	9,68 %

Отримані результати демонструють основні характеристики температурних показників і можуть бути використані для подальшого аналізу.

Контрольні запитання

1. Що таке описова статистика, і для чого її застосовують у екології?
2. Яка різниця між середнім значенням, модою та медіаною?
3. Як інтерпретувати коефіцієнт варіації?
4. Які висновки можна зробити на основі гістограми та Boxplot?

ПРАКТИЧНА РОБОТА №3: ПРОВЕДЕННЯ Т-ТЕСТУ ТА АНАЛІЗУ χ^2

Мета роботи: навчитися використовувати Т-тест для аналізу впливу екологічних факторів на різні змінні та критерій χ^2 для оцінки залежностей між категоріальними екологічними змінними. Показати, як ці статистичні методи застосовуються для вирішення реальних екологічних завдань.

Основні поняття

Тест хі-квадрат, також відомий як тест χ^2 , оцінює зв'язок між двома категоріальними змінними. Тест хі-квадрат є потужним інструментом для аналізу категоріальних даних. Це допомагає нам визначити, чи існує значний зв'язок між двома категоріальними змінними. Ось кілька ключових моментів, які слід враховувати. Статистика хі-квадрат вимірює розбіжність між спостереженими та очікуваними частотами в таблиці непередбачених обставин.

Критерій Хі-квадрат та Т-критерій є важливими інструментами у наборі статистичних інструментів. Незалежно від того, чи аналізуєте ви відповіді на опитування, дані клінічних випробувань чи дослідження ринку, розуміння цих тестів дозволить вам приймати обґрунтовані рішення.

Він кількісно визначає, наскільки наші спостережувані дані відхиляються від цього, що очікували при нульовій гіпотезі (т. е. відсутність зв'язку між змінними).

Завдання роботи

1. Виконати Т-тест для оцінки впливу одного екологічного фактора на середнє значення іншого.
2. Провести аналіз χ^2 для оцінки залежності між категоріальними екологічними змінними.
3. Інтерпретувати результати статистичних тестів у контексті екологічного аналізу.

Хід виконання роботи

1. Проведення Т-тесту

1.1. Мета тесту:

Перевірити, чи впливає рівень забруднення на середній рівень рослинного покриття в міських і сільських регіонах.

1.2. Гіпотези:

- H_0 : Середні значення рослинного покриття в міських і сільських регіонах однакові.
- H_a : Середні значення рослинного покриття в міських і сільських регіонах відрізняються.

1.3. Розрахунок середніх значень:

- Міський регіон:

$$\bar{x}_{\text{Urban}} = \frac{30 + 25 + 28}{3} = 27.67 \%$$

- Сільський регіон:

$$\bar{x}_{\text{Rural}} = \frac{70 + 75 + 68}{3} = 71.00 \%$$

1.4. Формула для Т-тесту:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Де:

\bar{x}_1, \bar{x}_2 – середні значення;

s_1^2, s_2^2 – дисперсії;

n_1, n_2 – кількість спостережень.

1.5. Використання програмного забезпечення:

- **R:**

R

Копіювати Редагувати

```
urban <- c(30, 25, 28)
rural <- c(70, 75, 68)
t.test(urban, rural, var.equal = TRUE)
```

- **Python:**

python

Копіювати Редагувати

```
from scipy.stats import ttest_ind
urban = [30, 25, 28]
rural = [70, 75, 68]
t_stat, p_value = ttest_ind(urban, rural, equal_var=True)
print("T-statistic:", t_stat, "P-value:", p_value)
```

1.6. Інтерпретація результатів:

- Якщо p -значення < 0.05 , відхиляємо H_0 . Різниця між середніми значуща.

- Якщо p -значення ≥ 0.05 , приймаємо H_0 . Різниця незначуща.

2. Проведення аналізу χ^2

2.1. Мета тесту:

Перевірити, чи залежить тип рослинного покриву від регіону.

2.2. Гіпотези:

- H_0 : Тип рослинного покриву не залежить від регіону.
- H_a : Тип рослинного покриву залежить від регіону.

2.3. Побудова таблиці спостережень:

Region	Sparse	Dense
Urban	3	0
Rural	0	3

2.4. Формула для χ^2 :

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Де O_{ij} – спостережувані значення, E_{ij} – очікувані значення.

2.5. Використання програмного забезпечення:

• R:

```
table <- matrix(c(3, 0, 0, 3), nrow = 2)
chisq.test(table)
```

• Python:

```
python
from scipy.stats import chi2_contingency
table = [[3, 0], [0, 3]]
chi2, p, dof, expected = chi2_contingency(table)
print("Chi-square:", chi2, "P-value:", p)
```

2.6. Інтерпретація результатів:

- Якщо p -значення < 0.05 , відхиляємо H_0 . Зв'язок між змінними є значущим.
- Якщо p -значення ≥ 0.05 , приймаємо H_0 . Зв'язок між змінними відсутній.

Вихідні дані

Датасет, що містить дані про вплив забруднення на стан рослинності в двох регіонах (міському та сільському).

Region	Pollutant_Level ($\mu\text{g}/\text{m}^3$)	Vegetation_Cover (%)	Vegetation_Type
Urban	55	30	Sparse
Urban	60	25	Sparse
Urban	58	28	Sparse
Rural	30	70	Dense
Rural	25	75	Dense
Rural	28	68	Dense

Звіт про виконання роботи

1. Мета та завдання роботи.
2. Таблиця вихідних даних.
3. Результати Т-тесту (з середніми значеннями та p -значенням).
4. Результати χ^2 (з таблицею спостережень та p -значенням).
5. Висновки та відповіді на контрольні запитання.

Контрольні запитання

1. Що таке Т-тест і коли його застосовують?
2. Як інтерпретувати результати критерію χ^2 ?
3. Чому важливо оцінювати вплив екологічних факторів на біологічні параметри?
4. У яких випадках можна відмовитися від нульової гіпотези?

ПРАКТИЧНА РОБОТА №4. НЕЛІНІЙНИЙ КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Мета: набути навичок у проведенні кореляційного аналізу криволінійної залежностей однієї величини від іншої під час вивчення екологічних систем. Повторити основні прийоми роботи з функціями Excel. Навчитися обчислювати коефіцієнт рангової кореляції, перевіряти його значимість і надійність за допомогою функцій Excel.

Хід роботи:

Приклад 1. Дано залежність між розмірами частинок у мкм (Y) та вмістом кальцію у % (X).

Розмір частинок, мкм	250	125	82	155	78	255	104	65	104	77	118
Вміст кальцію, %	23	15	22	8	12	45	21	32	11	18	25

Використовуючи можливості табличного процесору Excel:

- 1) обчислити коефіцієнт кореляції, перевірити його значимість і надійність;
- 2) обчислити параметри прямого і оберненого прогнозів;
- 3) побудувати графіки прогнозів і визначити кут між ними;
- 4) на основі отриманих результатів зробити висновки.

Крок 1. Вводимо вихідні дані в електронну таблицю Excel. Оскільки вихідна модель є нелінійною (рис. 1.) то нам необхідно знайти ранги.

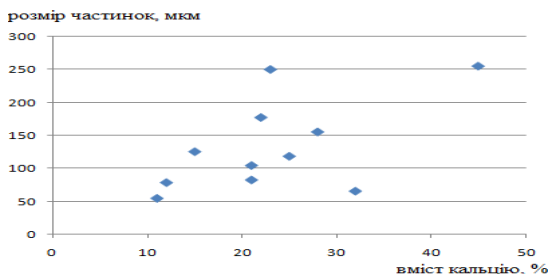


Рис. 1. Нелінійна залежність двох величин

Крок 2. Шукаємо ранги. Для цього використовуємо вікно «мастер функцій». Вибираємо функцію «РАНГ» і натискаємо «ОК» (рис.2.)

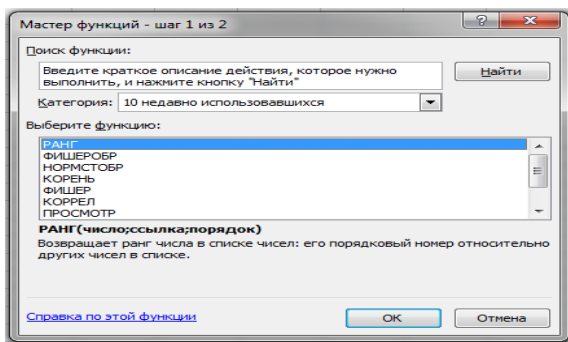


Рис. 2. Діалогове вікно

Вводимо дані (рис. 6.3.) для значень X та Y (для кожного ранги розраховуємо окремо).

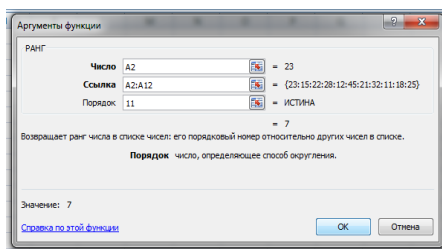


Рис. 3. Діалогове вікно «Аргументы функций» Отримуємо результат (табл.1.).

Таблица 1.

Таблица з вихідними даними та розрахованими рангами

	A	B	C	D
1	X	Y	Ранг X	Ранг Y
2	23	250	7	10
3	15	125	3	7
4	21	82	3	4
5	28	155	6	6
6	12	78	2	3
7	45	255	6	6
8	21	104	2	3
9	32	65	4	2
10	11	54	1	1
11	22	177	1	2
12	25	118	1	1

Крок 3. Розраховуємо коефіцієнт рангової кореляції Спірмена. Для

цього у вікні «функції» вибираємо «Вставити функцію» та у діалоговому вікні вибираємо функцію «КОРРЕЛ». Вводимо масиви (рис. 4.)

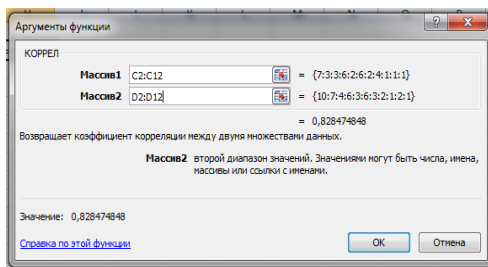


Рис. 4. Діалогове вікно

Для порівняння розраховуємо також коефіцієнт кореляції Пірсона. Результати представлено на рис. 5.

Таблиця 5.

Результати розрахунків

	A	B	C	D	E	F	G	H
1	X	Y	Ранг X	Ранг Y				
2	23	250	7	10		КОЕФІЦІЄНТ КОРЕЛЯЦІЇ СПІРМЕНА		0,828475
3	15	125	3	7		КОЕФІЦІЄНТ КОРЕЛЯЦІЇ ПІРСОНА		0,577281
4	21	82	3	4				
5	28	155	6	6				
6	12	78	2	3				
7	45	255	6	6				
8	21	104	2	3				
9	32	65	4	2				
10	11	54	1	1				
11	22	177	1	2				
12	25	118	1	1				

Аналізуючи дані рис. 5. відмічаємо, що коефіцієнт кореляції Спірмена (рівний 0,83) значено відрізняється від коефіцієнта кореляції Пірсона (0,58), а отже маємо справу з нелінійним характером взаємозв'язку між змінними.

Завдання для індивідуальної роботи:

Варіант 1

Встановити кореляційний зв'язок між двома змінними: X – висота труби, м; Y - об'єм викидів газоподібних сполук за добу, т.

X	25	15	82	100	28	37	16	32	78	52	34	65	21	61	91	45	71
Y	31	105	14	23	16	2	16	11	13	18	45	28	12	84	1	26	15

Варіант 2

Встановити кореляційний зв'язок між двома змінними: X – об'єм розчину, дм^3 ; Y – концентрація поліфосфатів, мг/дм^3 .

X	2,5	1,5	8,2	9,6	4,8	3,7	1,6	3,2	7,8	5,2	3,4	4,4	5,8	6,2	7,2
Y	0,1	10,5	1,4	0,3	0,6	0,3	1,6	1,7	7,3	1,8	4,5	0,5	3,5	7,3	8,3

Варіант 3

Встановити кореляційний зв'язок між двома змінними: X – урожайність, кг; Y – маса мінеральних добрив, кг.

X	215	115	182	110	328	307	106	321	788	521	348	224	845
Y	31	95	34	38	16	21	116	131	132	188	405	212	34

Варіант 4

Встановити кореляційний зв'язок між двома змінними: X – об'єм розчину, дм^3 ; Y – концентрація сульфат-іонів, мг/дм^3 .

X	2,1	2,5	3,2	3,1	4,8	3,7	3,6	3,2	2,8	5,2	3,4	4,4	2,2	2,2	3,2
Y	1,1	1,5	2,4	2,3	0,5	5,3	4,6	1,2	1,3	1,2	4,5	0,5	2,5	1,3	1,8

Варіант 5

Встановити кореляційний зв'язок між двома змінними: X – загальна площа лісу, тис. га; Y – частка території з ураженими деревами, %.

X	50	13	80	10	20	30	16	33	70	51	34	55	25	65	95	45	75
Y	5	10	30	10	20	10	40	16	22	15	12	17	24	19	15	28	51

Варіант 6

Встановити кореляційний зв'язок між двома змінними: X – об'єм розчину, дм^3 ; Y – маса сухого залишку, мг.

X	31,5	11,5	28,2	18,6	14,8	23,7	41,6	43,2	27,8	55,2	33,4	24,4
Y	11	23	14	34	23	14	17	24	27	28	15	18

Варіант 7

Встановити кореляційний зв'язок між двома змінними: X – загальна площа лісу, тис. га; Y – частка пошкодженої території лісу, %.

X	70	33	82	16	23	38	86	63	75	58	40	53	22	85	91	47	75
Y	15	71	30	20	40	30	14	60	12	50	21	52	43	19	11	25	17

Питання для самоконтролю:

1. Вимоги до вихідних даних для кореляційного аналізу.
2. Емпіричний коефіцієнт кореляції.
3. Рангові коефіцієнти кореляції.
4. Похибка та достовірність коефіцієнта кореляції.
5. Застосування нелінійних кореляційних методів аналізу в екології.

**ПРАКТИЧНА РОБОТА №5.
ЛІНІЙНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ**

Мета: сформувати вміння проводити необхідні розрахунки для визначення лінійного взаємозв'язку між двома змінними величинами, їх математичний вираз, побудови електронних таблиць та графіків для наочного представлення результатів досліджень.

Хід роботи:

Приклад №1. У результаті проведення експериментального дослідження отримано показники x (об'єм стічної води, дм^3) та y (вміст іонів кальцію, $\text{мг}/\text{дм}^3$) (табл. 1). Позначимо незалежну ознаку як x , а залежну – як y . Проаналізуємо вихідні дані за таким алгоритмом:

- ✓ для побудови кореляційного поля у таблиці Microsoft Office Excel (стовбці А і В) вносимо значення змінних x і y ;
- ✓ проводимо розрахунки за формулами, наведеними в табл. 1.;
- ✓ підбираємо рівняння лінійної регресії, розраховуємо коефіцієнти регресії та кореляції, перевіряємо значущість за допомогою критерію Стьюдента та аналізуємо тісноту кореляційного зв'язку між досліджуваними ознаками;
- ✓ будуюмо графік залежності між ознаками y і x та порівнюємо його з експериментальними даними. Отже, вводимо до таблиці експериментальні значення x та y ($N = 24$, де N -число експериментальних даних) і будуюмо кореляційне поле (табл. 1).

№	A	B	C	D	E	F	G	H	I	J	K	L
1	x	y	x*(ср.)	yy(ср.)	(x*(ср.)) ²	(yy(ср.)) ²	x ²	y ²	xy	(x*(ср.))y-y(ср.)	yx+xb	rs(y-b)/h
2	104,47	3,41	-26,1	-2,45	681,21	6,0025	10913,98	11,6381	356,2427	61,9485	3,26983	105,8838
3	123,03	5,04	-5,54	-0,82	30,6916	0,6724	15632,5	25,4016	630,1512	4,5428	5,30527	122,3485
4	126,93	5,41	-3,64	-0,45	13,2496	0,2025	16111,22	29,2681	686,6913	1,638	5,49337	126,0559
5	123,81	4,79	-6,76	-1,07	45,6976	1,1449	15328,92	22,9441	593,0499	7,2332	5,18449	119,8232
6	141,14	6,91	10,57	1,05	111,7249	1,1025	19920,5	47,7481	975,2774	11,0985	6,90016	141,2374
7	150,86	7,82	20,29	1,96	411,6841	3,8416	22758,74	61,1524	1179,725	39,7684	7,86244	150,4293
8	128,62	5,86	-1,95	0	3,8025	0	16543,1	34,3396	753,7132	0	5,66068	130,6313
9	133,66	6,08	3,09	0,22	9,5481	0,0484	17865	36,9664	812,6528	0,6798	6,15964	132,8535
10	132,81	6,01	2,24	0,15	5,0176	0,0225	17638,5	36,1201	798,1881	0,336	6,07549	132,1465
11	122,26	4,86	-8,31	-1	69,0561	1	14947,51	23,6196	594,1836	8,31	5,03104	120,5303
12	107,92	3,68	-22,65	-2,18	513,0225	4,7534	11646,73	13,5424	397,3456	49,377	3,61136	108,6111
13	103,36	3,29	-27,21	-2,47	740,3841	6,1009	10683,29	11,4921	350,3904	67,2067	3,15994	105,6818
14	137,54	6,21	6,97	0,35	48,5809	0,1225	18917,25	38,5641	854,1234	2,4395	6,54376	134,1667
15	121,18	4,87	-9,39	-0,99	88,1721	0,9801	14684,59	23,7169	590,1466	9,2561	4,92412	120,6313
16	145,79	7,61	15,22	1,75	231,6484	3,0625	21254,72	57,9121	1109,462	26,635	7,36051	148,3081
17	110,52	4,15	-20,05	-1,71	402,0025	2,9241	12214,67	17,2225	458,658	34,2855	3,86878	113,3586
18	160,04	8,72	29,47	2,86	888,4809	8,1796	25612,8	76,0384	1395,549	84,2842	8,77126	159,5202
19	159,08	8,55	28,51	2,69	812,8201	7,2361	25306,45	73,1025	1360,134	76,6919	8,67622	157,803
20	130,05	5,73	-0,52	-0,13	0,2704	0,0169	16913	32,8329	745,1865	0,0676	5,80225	129,3182
21	128,05	5,82	-2,52	-0,04	6,3504	0,0016	16396,8	33,8724	745,251	0,1008	5,60425	130,2273
22	126,94	5,05	-3,63	-0,81	13,1769	0,6561	16113,76	25,5025	641,047	2,9403	5,49436	122,4695
23	111,96	4,11	-18,61	-1,75	346,3321	3,0625	12335,04	16,8921	460,1566	32,5675	4,01134	112,9545
24	143,29	7,51	12,72	1,65	161,7984	2,7225	20532,02	56,4001	1076,108	20,988	7,11301	147,296
25	156,46	9,04	27,89	3,18	777,8521	10,1124	25109,57	81,7216	1432,478	88,6902	8,61444	162,7325
26	сума	сума	сума	сума	сума	сума	сума	сума	сума	сума	сума	сума

Рис. 1. Електронна таблиця експериментальних даних (стовбці А, В) і результати розрахунків (стовбці С - L).

Із розташування точок видно, що вони формують певну залежність, яка вірогідно має лінійний характер. Рівняння лінійної регресії у загальному виражаємо $y=kx+b$

Розрахунки проводимо наступним чином. У вікні «данные» вибираємо

«Анализ данных». В меню «инструменты анализа» вибираємо «Регрессия» і натискаємо «ок» (рис. 1.).

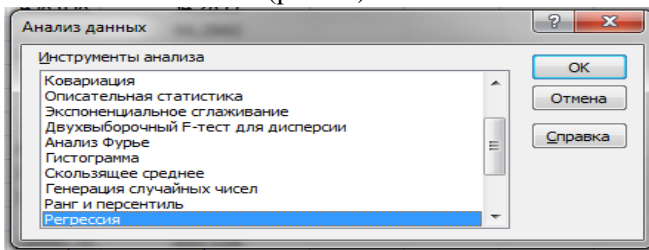


Рис.1. Вікно аналізу даних.

Результати розрахунків отримаємо у наступній таблиці 2. та 3.

Таблиця 2.

Результати розрахунків

	A	B	C	D	E
1	Вывод итогов				
2					
3	<i>Регрессионная статистика</i>				
4	Множественный R	0,9900822			
5	R-квадрат	0,9802628			
6	Нормированный R-квадрат	0,9793657			
7	Стандартная ошибка	0,2395582			
8	Наблюдения	24			
9					
10	Дисперсионный анализ				
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
12	Регрессия	1	62,70495645	62,70495645	1092,646347
13	Остаток	22	1,26253938	0,057388154	<i>Значимость F</i>
14	Итого	23	63,96749583		3,0054E-20
15					
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
17	Y-пересечение	-7,072513417	0,394271432	-17,93818383	1,27794E-14
18	Переменная X 1	0,099040556	0,002996217	33,05520152	3,0054E-20
19					
20		<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
21		-7,890182316	6,254844517	-7,890182316	-6,254844517
22		0,092826782	0,105254329	0,092826782	0,105254329

Аналізуючи дані таблиці 2. зазначимо, що *MS Excel* використовує наступні позначення:

SS - сума квадратів

- *df* - ступені вільності,
- *MS* - середній квадрат (дисперсія),
- *F* - F-статистика Фішера (фактичне значення),
- *p-значення* - значимість дисперсійного аналізу по кожному фактору окремо;
- *F-критичне* - критичне значення F-статистики при $p=0,05$.

Таблиця 3.

Результати обчислень

	A	B	C	D	E	F
20	Вывод остатка					Вывод вероятности
21						
22	<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>		<i>Перцентиль</i>	<i>Y</i>
23	1	3,274253429	0,135746571		2,083333333	3,39
24	2	5,310527253	-0,270527253		6,25	3,41
25	3	5,498704309	-0,088704309		10,416666667	3,68
26	4	5,189697775	-0,399697775		14,583333333	4,11
27	5	6,90670604	0,003929396		18,75	4,15
28	6	7,868744805	-0,048744805		22,916666667	4,79
29	7	5,666082848	0,193917152		27,083333333	4,86
30	8	6,165247248	-0,085247248		31,25	4,87
31	9	6,081062776	-0,071062776		35,416666667	5,04
32	10	5,036184914	-0,176184914		39,583333333	5,05
33	11	3,615943346	0,064056654		43,75	5,41
34	12	3,164318413	0,225681587		47,916666667	5,73
35	13	6,549524604	-0,339524604		52,083333333	5,82
36	14	4,929221114	-0,059221114		56,25	5,86
37	15	7,366609188	0,243390812		60,416666667	6,01
38	16	3,873448791	0,276551209		64,583333333	6,08
39	17	8,777937106	-0,057937106		68,75	6,21
40	18	8,682858172	-0,132858172		72,916666667	6,91
41	19	5,807710842	-0,077710842		77,083333333	7,51
42	20	5,609629731	0,210370269		81,25	7,61
43	21	5,499694714	-0,449694714		85,416666667	7,82
44	22	4,016067191	0,093932809		89,583333333	8,55
45	23	7,119007799	0,390992201		93,75	8,72

У результаті проведення розрахунків отримуємо такі коефіцієнти

регресії k і b та коефіцієнт кореляції R : $k = 0,099$; $b = -7,0725$; $R = 0,99$
 Оскільки значення R наближається до 1, можемо зробити висновки, що існує дуже тісний кореляційний зв'язок між незалежною ознакою x та залежною ознакою y .

Отримане рівняння регресії має вигляд: $y = 0,099 * x - 7,0725$.
 Графічний вигляд даної лінійної залежності має вид (рис.2.).

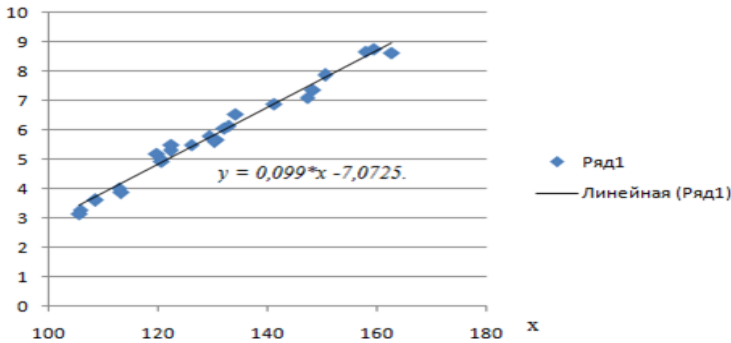


Рис.2. Графічний вигляд кореляційного поля.

Для перевірки розрахунків розв'язуємо обернену задачу – результат розрахунку оберненої задачі наведено в таб.4.

	A	B		A	B
1	y	y'	1	y	y'
2	3,41	3,27	14	6,21	6,55
3	5,04	5,31	15	4,87	4,93
4	5,41	5,50	16	7,61	7,37
5	4,79	5,19	17	4,15	3,87
6	6,91	6,91	18	8,72	8,78
7	7,82	7,87	19	8,55	8,68
8	5,86	5,67	20	5,73	5,81
9	6,08	6,17	21	5,82	5,61
10	6,01	6,08	22	5,05	5,50
11	4,86	5,04	23	4,11	4,02
12	3,68	3,62	24	7,51	7,12
13	3,39	3,16	25	9,04	8,62

Рис.6. Результати перевірки правильності розрахунків: y – експериментальна величина; y' - розрахункова величина.

Як можна побачити з даних табл. 4., значних розбіжностей між експериментальними і розрахованими даними не спостерігається.

Завдання для індивідуальної роботи:

Завдання 1.

Перевірити наявність кореляційного зв'язку між величинами, побудувати графік функції $y = f(x)$ і запропонувати рівняння регресії. Розрахувати коефіцієнт кореляції, оцінити його значущість. Надати таблицю експериментальними.

Варіант № 1. Отримано такі дані щодо внесення мінеральних добрив у ґрунти x (кг) та врожайності культур y (кг):

X	0,8	0,6	0,5	0,4	0,7	0,7	0,8	1,0	1,3	1,4	1,2	1,5	1,8	2,0	2,2	2,3
Y	21	14	16	11	18	19	20	25	28	30	25	33	41	48	51	53

Варіант № 2. Отримано такі дані щодо кількості викинутого оксиду Нітрогену x (мг/дм³) та відстані, на якій його концентрація досягне нормованих значень y (км):

X	0,02	0,03	0,05	0,04	0,07	0,03	0,08	1,04	1,03	1,07	1,12	1,15
Y	4	5	6	5	8	4	10	13	12	17	18	21

Варіант №3. Отримано такі дані щодо внесення фунгіцидів у ґрунти x (кг) та врожайності культур y (кг):

X	1,7	1,4	1,5	1,4	1,9	1,7	1,8	1,0	1,3	1,4	1,2	1,5	1,8	2,0
Y	212	208	216	211	268	221	230	185	202	213	215	213	241	237

Варіант № 4. Отримано такі дані щодо кількості викинутого оксиду Карбону x (мг/дм³) та відстані, на якій його концентрація досягне нормованих значень y (км):

X	0,12	0,03	0,25	0,42	0,71	0,23	0,18	0,04	0,83	0,47	0,62	0,75
Y	4	1	1	14	22	8	5	2	31	17	18	25

Варіант № 5. Отримано такі дані щодо внесення гербіцидів у ґрунти x (кг) та врожайності культур y (кг):

X	5	5,2	7	8	11	13	17	14	24	21	28	22	32
Y	282	284	378	421	688	723	820	725	1028	1003	1425	1241	1432

Варіант № 6. Отримано такі дані щодо кількості викинутого оксиду Сульфуру x (мг/дм³) та відстані, на якій його концентрація досягне нормованих значень y (км):

X	0,21	0,35	0,51	0,41	0,72	0,38	0,83	0,64	0,73	0,47	0,52	0,95
Y	8	14	26	35	48	14	53	43	51	37	37	61

Варіант № 7. Отримано такі дані щодо внесення мінеральних добрив у ґрунти x (кг) та врожайності культур y (кг):

X	1,8	1,6	2,0	1,4	1,7	1,9	1,8	2,2	2,3	2,6	2,2	3,0	3,8	4,0	4,7	5,3
Y	21	14	21	11	18	20	19	25	28	30	25	33	41	48	51	53

Питання для самоконтролю:

1. Поняття про регресію, регресійне поле.
2. Лінійні моделі та обмеження їх застосування.
3. Дисперсія.
4. Середньоквадратичне відхилення.
5. Похідна.
6. Загальний вигляд рівняння регресії.
7. Методологія регресійного аналізу.
8. Основні завдання регресійного аналізу.
9. Застосування регресійних моделей в екології.
10. Види регресії.
11. Переваги та недоліки лінійного регресійного аналізу.

ПРАКТИЧНА РОБОТА № 6.

ДВОХФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ ІЗ ДАНИМИ БЕЗ ПОВТОРЕНЬ

Мета: сформувати вміння виявляти фактори (контрольовані чинники), які впливають на кінцевий результат досліджень. Набути навичок у процесі проведення двохфакторного дисперсійного аналізу в Excel.

Хід роботи

Приклад. Групи чоловіків та жінок запропонували послухати звук різної гучності. Час відповіді фіксували у мілісекундах. Необхідно визначити, чи впливає на реакцію стать та гучність. Отримані дані занесені до таблиці (табл. 1.) .

Таблиця 1.

Вихідні дані для аналізу

1	A	B	C	D	E	F
2	гучність звуку, дБ					
3	стать	10	20	30	40	50
4	жінки	272	254	233	202	182
5	чоловіки	305	258	247	238	225

Для вирішення задачі в *MS Excel*:

1. Виконуємо команду «Анализ данных» із меню «Данные».
2. У діалоговому вікні (рис. 1.) необхідно вибрати метод «Двухфакторный дисперсионный анализ без повторений» та натиснути кнопку [OK].

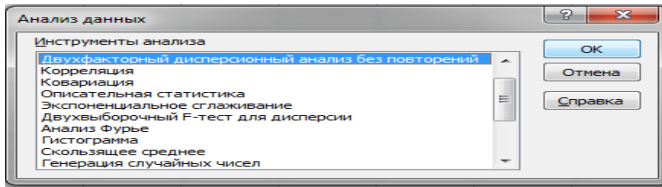


Рис. 1. Вікно «Анализ данных»

Отримуємо результат (табл. 2.).

Таблиця 2.

Результати аналізу

1	A	B	C	D	E	F	G
2	Двухфакторный дисперсионный анализ без повторений						
3	ИТОГИ						
4		Счет	Сумма	Среднее	Дисперсия		
5	Строка 1	5	1143	228,6	1356,8		
6	Строка 2	5	1273	254,6	940,3		
7	Столбец 1	2	577	288,5	544,5		
8	Столбец 2	2	512	256	8		
9	Столбец 3	2	480	240	98		
10	Столбец 4	2	440	220	648		
11	Столбец 5	2	407	203,5	924,5		
12							
13							
14	Дисперсионный анализ						
15	Источник вариации						
16		SS	df	MS	F	P-Значение	F критическое
17	Строки	1690	1	1690	12,6829268	0,023558366	7,708647421
18	Столбцы	8655,4	4	2163,85	16,2390244	0,009704373	6,388232909
19	Погрешность	533	4	133,25			
20	Итого	10878,4	9				

Примітка: *SS* - сума квадратів, *df* - ступені вільності, *MS* - середній квадрат (дисперсія), *F* - F-статистика Фішера (фактичне значення), *p-значення* - значимість дисперсійного аналізу по кожному фактору окремо; *F-критичне* - критичне значення F- статистики при $p=0,05$.

Аналізуючи дані табл.12.2. можемо відмітити, що оскільки критерій

F- статистики для фактору «стать» більше за критичний рівень F-розподілу (стовбець «F-критическоне»), то даний фактор має вплив на досліджуваний параметр (у нашому прикладі це час реакції на звук).

Сума квадратів, обумовлена впливом першого фактора (гучність звуку), дорівнює 1690. Сума квадратів, обумовлена впливом другого чинника (стать), дорівнює 8655,4. Залишкова, внутрішньогрупова сума квадратів (похибка) дорівнює 533. Залишкова, внутрішньогрупова дисперсія (похибка) дорівнює 133,25.

Завдання для індивідуальної роботи:

Використовуючи програмне забезпечення MS Excel провести двохфакторний дисперсійний аналіз із даними без повторень.

Результати представити у вигляді таблиці. Зробити необхідні висновки *Варіант 1.*

Маса листка, гр.							
Висота листка, см	Ширина листка, см						
	a=2	a=3	a=4	a=5	a=6	a=7	a=8
h=3	1	2	3	4	5	6	7
h=4	2	3	4	5	6	7	8
h=5	3	5	8	10	11	13	15
h=6	4	7	8	11	14	15	17
h=7	5	10	11	13	16	18	21
h=8	5	11	13	14	18	22	24

Варіант 2.

Вік дітей	Гучність звуку, дБ						
	5	10	15	20	25	30	35
до 3 років	238	212	194	183	175	167	158
від 3 до 6 років	257	234	224	211	207	195	184
від 6 до 12 років	264	258	243	231	228	217	198
від 12 до 16 років	266	251	240	230	226	224	204
від 16 до 18 років	282	273	264	253	241	227	218

Варіант 3.

		Маса листка, гр.					
Висота листка, см	Ширина листка, см						
	a=2,5	a=3	a=3,5	a=4	a=4,5	a=5	a=5,5
h=2,3	1,8	2,1	3,2	4,4	5,6	6,2	7,6
h=3,4	2,1	3,6	4,8	5,3	6,8	7,0	8,3
h=3,7	3,5	5,8	8,3	10,2	11,5	13,4	15,5
h=4,2	4,2	7,2	8,7	11,5	14,8	15,5	17,7
h=4,7	5,1	10,4	11,3	13,2	16,9	18,3	21,3
h=5,2	5,7	11,1	13,8	14,3	18,4	22,7	24,3

Варіант 4.

		Гучність звуку, дБ					
Вік дорослих	5	10	15	20	25	30	35
від 20 до 30 років	267	264	257	241	237	225	218
від 30 до 40 років	271	265	254	248	238	228	205
від 40 до 50 років	276	271	251	250	245	236	217
від 50 до 60 років	280	277	263	251	253	247	221
від 60 до 70 років	289	282	271	258	261	256	234

Питання для самоконтролю:

1. Способи представлення інформації про результати дослідження: графічні, схематичні, табличні.
2. Корисність одержаних результатів.
3. Використання одного показника при виборі рішення, що пов'язане з ризиком. Дерево рішень.
4. Приклад вибору дій, необхідних для підвищення рівня споживання природних ресурсів без суттєвого забруднення навколишнього середовища.
5. Використання одного показника при виборі рішення, що пов'язаний з ризиком .
6. Взаємозв'язок між дискримінантними змінними і дискримінантною функцією.
7. Теоретичні та практичні аспекти дискримінантного аналізу в екології.
8. Дослідження структури і функціонування екосистем

ПРАКТИЧНА РОБОТА №7: АНАЛІЗ ЧАСОВИХ РЯДІВ ТА ПРОГНОЗУВАННЯ ЕКОЛОГІЧНИХ ЗМІН

Мета роботи: ознайомитися із методами аналізу часових рядів для оцінки змін екологічних показників у часі та побудови прогнозів. Навчити інтерпретувати результати аналізу та використовувати їх для екологічного моніторингу та планування заходів з охорони довкілля.

Завдання роботи

1. Провести аналіз трендів у часовому ряді екологічних даних.
2. Виконати декомпозицію часового ряду на трендову, сезонну та залишкову компоненти.
3. Побудувати прогноз змін екологічного показника на основі моделі ARIMA або експоненціального згладжування.
4. Інтерпретувати отримані результати в контексті змін довкілля.

Хід виконання роботи

1. Аналіз трендів у часовому ряді

1.1. Візуалізація даних

- Побудуйте графік рівня забруднення у часі.

○ R:

```
data <- read.csv("time_series_environment.csv")
plot(data$Date, data$Pollutant_Level, type = "l", col = "blue",
      xlab = "Дата", ylab = "Рівень забруднення (µg/m³)", main = "Часовий ряд")
```

○ Python:

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("time_series_environment.csv", parse_dates=["Date"])
plt.plot(data["Date"], data["Pollutant_Level"], color="blue")
plt.title("Часовий ряд: Рівень забруднення")
plt.xlabel("Дата")
plt.ylabel("Рівень забруднення (µg/m³)")
plt.show()
```

1.2. Виявлення тренду

- Використайте метод ковзного середнього для згладжування даних.

○ R:

```
library(zoo)
data$Moving_Avg <- rollmean(data$Pollutant_Level, k = 3, fill = NA)
lines(data$Date, data$Moving_Avg, col = "red")
```

○ Python:

```
data["Moving_Avg"] = data["Pollutant_Level"].rolling(window=3).mean()
plt.plot(data["Date"], data["Moving_Avg"], color="red", label="Ковзне середнє")
plt.legend()
plt.show()
```

2. Декомпозиція часового ряду

2.1. Мета:

Розділити часовий ряд на:

- Трендову компоненту.
- Сезонну компоненту.
- Залишкову компоненту (шум).

2.2. Виконання декомпозиції:

• R:

```
library(stats)
ts_data <- ts(data$Pollutant_Level, start = c(2024, 1), frequency = 12)
decomposed <- decompose(ts_data)
plot(decomposed)
```

• Python:

```
from statsmodels.tsa.seasonal import seasonal_decompose

ts_data = data.set_index("Date")["Pollutant_Level"]
decomposition = seasonal_decompose(ts_data, model="additive", period=12)
decomposition.plot()
plt.show()
```

3. Прогнозування

3.1. Побудова моделі прогнозування

Використовуйте модель ARIMA або експоненціальне згладжування.

• R:

```
library(forecast)
model <- auto.arima(ts_data)
forecasted <- forecast(model, h = 12)
plot(forecasted)
```

- **Python:**

```
from statsmodels.tsa.arima.model import ARIMA

model = ARIMA(ts_data, order=(1, 1, 1))
fitted_model = model.fit()
forecast = fitted_model.forecast(steps=12)
plt.plot(ts_data, label="Історичні дані")
plt.plot(forecast, label="Прогноз", color="green")
plt.legend()
plt.show()
```

3.2. Інтерпретація прогнозу:

- Оцініть майбутню динаміку рівня забруднення на основі отриманих результатів.
- Проаналізуйте, чи спостерігаються тенденції до зростання або зниження забруднення.

Вихідні дані

Файл: time_series_environment.csv.

Date	Pollutant_Level ($\mu\text{g}/\text{m}^3$)	Temperature ($^{\circ}\text{C}$)
2024-01-01	45	12,5
2024-02-01	50	13,0
2024-03-01	55	15,0
2024-04-01	52	16,5
2024-05-01	60	18,0
2024-06-01	65	20,0
2024-07-01	70	22,5
2024-08-01	68	22,0
2024-09-01	65	20,0
2024-10-01	60	18,0
2024-11-01	55	15,0
2024-12-01	50	13,5

Звіт про виконання роботи

1. Мета та завдання роботи.
2. Вихідні дані (таблиця або графіки).

3. Результати аналізу трендів і декомпозиції (графіки та інтерпретація).
4. Результати прогнозу (графік та числові значення).
5. Висновки та відповіді на контрольні запитання.

Контрольні запитання

1. Які основні компоненти часового ряду?
2. Що таке ковзне середнє, і для чого його використовують?
3. Як можна інтерпретувати результати декомпозиції?
4. Чому ARIMA вважається однією з найбільш популярних моделей прогнозування?
5. Як результати аналізу часових рядів можуть бути використані в екології?

СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Глинський Я. М. Практикум з інформатики. Львів : СПД Глинський, 2010. 304 с.
2. Манн К. Б. Статистичні методи в екології. Київ : Наукова думка, 2020.
3. Данилко В. К. Екологічна статистика: водні ресурси : монографія. Київ, 2003. 368 с.
4. Тарасова В. В. Екологічна статистика (з блочно-модульною формою контролю знань) : підручник. Київ : Центр учбової літератури, 2008. 392 с.
5. Чекотовський Е.В. Основи статистики сільського господарства. Навч. посібник. К., 2001. 432 с.
6. Теорія статистики : навчальний посібник / М. К. Шапочка, О. М. Маценко. Суми : Університетська книга, 2014. 312 с.
7. Руденко В. М. Математична статистика : навч. посіб. К. : Центр учбової літератури, 2012. 304 с.
8. Методичні вказівки до виконання практичних робіт з навчальної дисципліни «Екологічна статистика» для здобувачів вищої освіти першого (бакалаврського) рівня та другого (магістерського) рівня усіх освітньо-професійних програм спеціальностей НУВГП денної і заочної форм навчання [Електронне видання] / Буднік З. М. Рівне : НУВГП, 2021. 36 с.
9. Ковалевський Г. В. Статистика : підручник. Х. : ХНАМГ, 2012. 445с.

10. Матковський С. О., Марець О. Р. Теорія статистики : навч. посібник. К. : Знання, 2010. 534 с.
11. Тринько Р. І., Стадник М. Є. Основи теоретичної і практичної статистики : навч. посібник. К. : Знання, 2011. 397 с.
12. Прикладна статистика : навч. посібник / В. О. Костюк; Харк. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. Харків : ХНУМГ ім. О. М. Бекетова, 2015. 191 с.
13. Інтегральні та комплексні оцінки стану навколишнього природного середовища: монографія / О. Г. Васенко та ін. Харків : НУГЗУ, 2015. 419 с.
14. Данилко В. К. Екологічна статистика України: здобутки і проблеми. *Статистика України*. № 1, 2002. С. 54–57.