

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО"  
ІНСТИТУТ ЕНЕРГОЗБЕРЕЖЕННЯ ТА  
ЕНЕРГОМЕНЕДЖМЕНТУ

**О. В. Кофанова, О. Є. Кофанов**

**МЕТОДИЧНІ ВКАЗІВКИ  
ДЛЯ САМОСТІЙНОЇ РОБОТИ  
З ДИСЦИПЛІНИ "КІЛЬКІСНИЙ  
ТА ЯКІСНИЙ АНАЛІЗ СТАНУ  
ЕКОЛОГІЧНИХ СИСТЕМ". ОБРОБКА  
РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТАЛЬНИХ  
ДОСЛІДЖЕНЬ**

Затверджено  
на засіданні Вченої Ради  
Інституту енергозбереження  
та енергоменеджменту  
КПІ ім. Ігоря Сікорського  
протокол № 2 від 25.09.2017 р.

Київ  
ІЕЕ КПІ ім. Ігоря Сікорського  
2018

УДК 502/504  
ББК 20.1  
К 31

Гриф надано Вченою радою ІЕЕ КПІ ім. Ігоря Сікорського  
(Протокол № 2 від 25.09.2017 р.)

Рецензенти: *О. М. Назаров*, доктор фізико-математичних наук, професор, провідний науковий співробітник Інституту фізики напівпровідників ім. В. Є. Лашкарьова НАН України;

*С. В. Зайченко*, доктор технічних наук, доцент кафедри електромеханічного обладнання енергоємних виробництв КПІ ім. Ігоря Сікорського.

Редактор: професор *К. К. Ткачук*, доктор технічних наук, завідувач кафедри інженерної екології КПІ ім. Ігоря Сікорського.

Кофанова О. В., Кофанов О. Є.

К 31 Методичні вказівки для самостійної роботи з дисципліни "Кількісний та якісний аналіз стану екологічних систем". Обробка результатів експериментальних досліджень [для підготовки докторів філософії за спеціальністю 101 "Екологія"] [Електронне видання] / О. В. Кофанова, О. Є. Кофанов. – К.: ІЕЕ КПІ ім. Ігоря Сікорського, 2018. – 41 с.

ISBN 000-000-000-0

У навчальному виданні викладено матеріал для підготовки докторів філософії зі спеціальності 101 "Екологія" до розв'язання комплексних завдань прогнозування стану навколишнього середовища методами кореляційного та регресійного аналізу. Видання створено із врахуванням сучасних потреб екологічних досліджень; особлива увага приділена питанням обробки результатів хімічних та екологічних досліджень стану літосфери, гідросфери й атмосфери, зокрема, при аналізі ступенів забруднення компонентів довкілля різноманітними поллютантами. Видання підвищує рівень забезпеченості педагогічного процесу підготовки докторів філософії з дисципліни "Кількісний та якісний аналіз стану екологічних систем" навчальним матеріалом та сприяє більш повному засвоєнню дисципліни.

УДК 502/504  
ББК 20.1  
ISBN 000-000-000-0  
© О. В. Кофанова, 2018  
© О. Є. Кофанов, 2018

## **З М І С Т**

ВСТУП	3
INTRODUCTION	5
РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ	7
1.1 Статистичний зв'язок між випадковими величинами	7
1.2 Оцінка значущості коефіцієнта кореляції	17
РОЗДІЛ 2 ТЕОРЕТИЧНІ ОСНОВИ РЕГРЕСІЙНОГО АНАЛІЗУ	19
РОЗДІЛ 3 АЛГОРИТМИ ТА ПРИКЛАДИ ВИКОНАННЯ РОЗРАХУНКІВ (ЛІНІЙНА РЕГРЕСІЯ)	24
РОЗДІЛ 4 ПРИКЛАДИ ЗАВДАНЬ ДО САМОСТІЙНОЇ РОБОТИ	31
ВИКОРИСТАНА ЛІТЕРАТУРА	37
ДОДАТКИ	39
ДОДАТОК А	40
ДОДАТОК Б	41

## В С Т У П

На сьогодні статистичні методи і різноманітні математичні моделі широко застосовуються для вивчення компонентів навколишнього природного середовища, контролю стану екологічних систем тощо. У навчальному виданні викладено матеріал для підготовки докторів філософії до розв'язання комплексних завдань прогнозування стану навколишнього середовища методами кореляційного і регресійного аналізу.

Видання "Методичних вказівок для самостійної роботи з дисципліни "Кількісний та якісний аналіз стану екологічних систем". Обробка результатів експериментальних досліджень" для підготовки докторів філософії є актуальним з огляду на проблеми навколишнього середовища й необхідність застосування "зелених" та енергозберігаючих технологій. Видання підвищує рівень забезпеченості матеріалом навчального процесу підготовки докторів філософії з дисципліни "Кількісний та якісний аналіз стану екологічних систем" та створене із врахуванням сучасних потреб проведення й обробки результатів хімічних і екологічних досліджень.

Екологічні явища формуються під впливом багатьох чинників, тому наявні взаємозв'язки між ознаками, як правило, легко визначати тільки у великих за обсягом системах. Для аналізу результатів хімічних і екологічних досліджень досить часто застосовують регресійний та кореляційний методи аналізу. Зокрема, кореляційний аналіз надає змогу розв'язати низку складних завдань екології та охорони навколишнього середовища і сприяє підвищенню точності й надійності математичних моделей.

Перевага кореляційного методу полягає також і в тому, що він надає змогу розв'язувати завдання, які не можна вирішити за допомогою інших методів аналізу, таких, наприклад, як відокремлення впливу кожного з багатьох чинників, що взаємопов'язано і взаємозумовлено діють на систему. Більш того, в екологічних дослідженнях необхідно визначати не стільки міру кореляції, скільки її форму і характер зміни однієї ознаки залежно від зміни іншої.

Під час вивчення курсу за допомогою "Методичних вказівок для самостійної роботи з дисципліни "Кількісний та якісний аналіз стану екологічних систем". Обробка результатів експериментальних досліджень" доктори філософії повинні оволодіти теоретичним матеріалом і практичними навичками застосування кореляційного аналізу під час проведення експериментів щодо визначення і контролю параметрів навколишнього середовища. Це надасть їм змогу виконувати якісну експертну оцінку отриманих результатів, робити коректні висновки та приймати вірні й екологічно безпечні управлінські рішення.

## INTRODUCTION

Currently, statistical methods and mathematical models are widely used to study components of the environment, for monitoring ecological systems state, etc. The study guide presents material for the postgraduate students' preparation in order to ensure their ability to solve complex tasks of forecasting the environment state using correlation and regression analysis methods.

The publication of the "Methodical guidelines for independent work on discipline "Quantitative and qualitative analysis of the ecological systems state". Processing results of the experimental researches" for the preparation of postgraduate students is relevant because of the environmental problems aggravation and the importance of the implementation of the "green" and energy-saving technologies. The publication enhances the material for the postgraduate students' training at the discipline "Quantitative and qualitative analysis of the ecological systems state" and was created taking into account the current needs of conducting and processing the results of chemical and environmental researches.

Environmental phenomena are formed because of the influence of many factors, so, in the most cases, the existing interrelationships between the features can be easily determined only in the large-scale systems. So, regression and correlation analysis methods are often used for analyzing the results of chemical and environmental studies. In particular, the correlation analysis enables to solve a number of complex environmental issues and improves the accuracy and reliability of the mathematical models.

The advantage of the correlation method is that it also allows scientists to solve tasks that can not be solved by other methods, for example, separating the influence of each of the factors that affect the particular ecosystem and are interconnected and interdependent. Moreover, in the environmental studies it is necessary to determine not only the degree of correlation but also its form and the peculiarities of the change of one feature, depending on the change of another.

During the course with the help of the "Methodical guidelines for independent work on discipline "Quantitative and qualitative analysis of the ecological systems state". Processing results of the experimental researches" postgraduate students will master the theoretical material and practical skills of the correlation analysis implementation during experiments on the determination and control of the environmental parameters. This will enable them to perform an expert assessment of the obtained results, make correct conclusions and well-thought-out environmentally friendly managerial decisions.

## РОЗДІЛ 1

### ТЕОРЕТИЧНІ ОСНОВИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ

#### 1.1 Статистичний зв'язок між випадковими величинами

У статистичному аналізі розрізняють такі види зв'язків між чинниками, що діють на систему (об'єкт), а саме:

- функціональні;
- стохастичні;
- статистичні.

*Функціональний зв'язок* – це зв'язок між змінними, при якому кожному значенню однієї величини відповідає чітко визначене значення іншої, тобто  $y = f(x_1, x_2, \dots, x_n)$ . Зазначимо, що дослідженнями зв'язків такого виду статистика не займається.

За наявності двох випадкових величин  $x$  і  $y$  часто виникає потреба у встановленні та врахуванні залежності між ними – так званого стохастичного зв'язку, що відповідає ситуації, за якої зміна значення однієї змінної спричинює зміну закону розподілу іншої. У випадку дискретних величин це означає, що певному значенню однієї змінної відповідає набір значень іншої і при цьому кожне значення має власну ймовірність реалізації.

*Статистичний зв'язок* передбачає, що значення однієї змінної варіюються в середньому в залежності від того, які значення приймає інша змінна. Часто статистичний зв'язок розглядають як функціональну залежність (функціональний зв'язок) з випадковою помилкою  $\varepsilon$ , зокрема:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon, \quad (1.1)$$

де  $f(x_1, x_2, \dots, x_n)$  – функція, що описує залежність  $y$  від сукупності незалежних змінних  $x_1, x_2, \dots, x_n$ .



Оскільки сума константи і випадкової величини є випадковою величиною, то значення  $y$ , розраховані за формулою (1.1), також будуть випадковими величинами.

На практиці для одержання за експериментальними даними математичних моделей, що описують поведінку певної характеристики в залежності від зміни декількох факторів, застосовують *регресійний аналіз*. При цьому екологічні процеси, об'єкти або природні явища розглядають як "чорний ящик" ("чорну скриню"), що має на "вході" незалежні змінні, якими можна керувати, а на "виході" – змінні, що характеризують досліджуваний процес.

Виокремлюють:

- контрольовані;
- керовані;
- некеровані та
- неконтрольовані фактори (чинники).

Зокрема, *некерованими факторами* називають незалежні змінні, значення яких можна вимірити, але керувати ними не можливо. *Неконтрольовані фактори* – це ті, що впливають на досліджуваний процес, але їх неможливо виміряти (або може бути невідомо про їх існування).

Основною ознакою наявності статистичного (кореляційного) зв'язку між випадковими величинами  $x$  та  $y$  є відмінність від нуля *коефіцієнту кореляції*  $r$ . За наявності кореляційного зв'язку між змінними необхідно виявити його характер, тобто визначити, як саме у середньому впливає зміна величини  $x$  на зміну величини  $y$ .

Отже, під *кореляційним аналізом* розуміють встановлення закономірностей між явищами і процесами, що залежать від багатьох, іноді невідомих параметрів. Використання методу кореляції і регресії дає змогу розв'язати такі завдання:

- встановити характер і тісноту зв'язку між досліджуваними величинами (явищами);

- визначити й кількісно виміряти ступінь впливу окремих факторів (чинників) та їх комплексної дії на певний об'єкт чи явище;
- на підставі кореляційної моделі розраховувати (прогнозувати) кількісні зміни у стані об'єкта чи явища, надавати об'єктивну оцінку цим змінам.

Таким чином, функція  $y = f(x)$  є *кореляційною*, якщо кожному значенню аргументу  $x$  відповідає статистичний ряд розподілу функції  $y$ . Оскільки кореляційні залежності характеризуються *ймовірнісними зв'язками*, встановлення залежностей між величинами  $x$  та  $y$  можливе лише тоді, коли були проведені статистичні вимірювання.

Для отримання статистично достовірної і надійної характеристики зв'язку у дослідженні (спостереженні) потрібно встановити форму зв'язку й знайти кількісні характеристики – обрати математичне рівняння (модель), яке найбільш повно відображає характер зв'язку між досліджуваними ознаками явища чи об'єкта.

Отже, під *формою кореляційного зв'язку* розуміється тип аналітичного рівняння, що виражає залежність між певними досліджуваними ознаками. Зв'язок між корелюючими ознаками за *напрямом* може бути:

- *прямим*, коли обидві ознаки змінюються в одному напрямі (зі збільшенням факторної ознаки зростає ознака результативна і навпаки);
- *оберненим*, коли обидві ознаки змінюються в різних напрямках.

За формою або аналітичним вираженням розрізняють:

- *прямолінійні зв'язки (або лінійні)*;
- *нелінійні (криволінійні) зв'язки*.

Залежно від кількості досліджуваних ознак кореляція може бути:

- *парною (простою)* і
- *множинною*.

Тобто якщо визначається зв'язок між двома ознаками (результативною і факторною), то таку кореляцію називають простою; а у випадку, коли

явище розглядається як результат одночасного впливу декількох чинників – множинною.

Встановлення форми зв'язку дає можливість знайти рівняння регресії, яке найбільш повно відображає характер взаємодії між залежною і незалежною ознаками. Проте сам метод кореляційного аналізу не встановлює наявності та/або причин виникнення зв'язків (причинно-наслідкового зв'язку) між досліджуваними величинами (явищами). Його призначенням є тільки встановлення наявності та кількісне описання зв'язку.

*Кореляційний аналіз можна поділити на 5 етапів:*

- встановлення наявності зв'язку між досліджуваними ознаками, для чого потрібно спочатку визначити та відібрати найістотніші чинники для аналізу;
- вивчення характеру зв'язку, його напрямку і форми; вибір математичного рівняння для його описання;
- розрахунок характеристик кореляційного зв'язку, а саме знаходження коефіцієнтів рівняння регресії і показників тісноти зв'язку;
- статистичне оцінювання характеристик (показників) зв'язку.

Таким чином, сутність кореляційного аналізу полягає у знаходженні *рівняння регресії* між випадковими величинами (так звана лінія найкращого наближення), а також в оцінці виду і тісноти зв'язку між ними, достовірності результатів вимірювань (спостережень) тощо.

Щоб попередньо оцінити наявність кореляційного зв'язку між випадковими величинами  $x$  та  $y$ , будують так зване *кореляційне поле*. Для цього результати випробувань (емпіричний ряд розподілу) наносять точками на координатну площину  $XOY$  (рис. 1.1). Множина таких точок й утворює кореляційне поле. За тіснотою групування експериментальних точок навколо прямої (кривої) лінії, а також за нахилом цієї лінії можна візуально оцінити наявність (чи відсутність) кореляційного зв'язку між досліджуваними величинами (явищами).

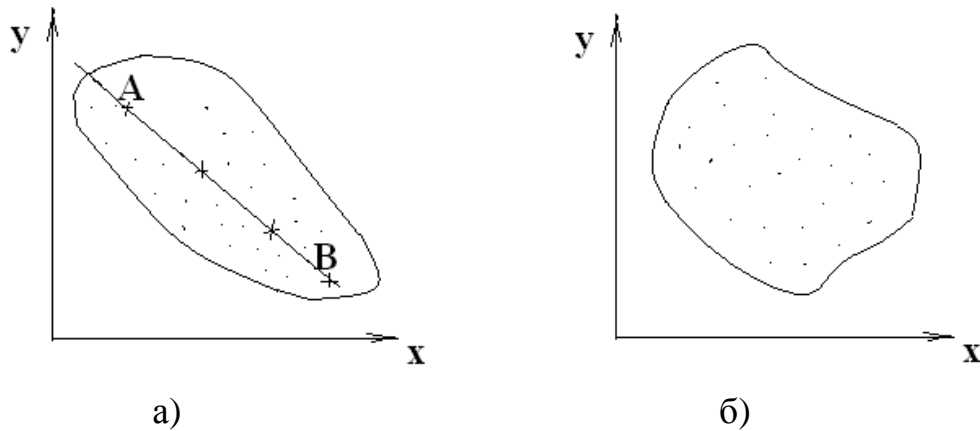


Рисунок 1.1 – Види кореляційних полів.

Наприклад, з рис. 1.1а) випливає, що експериментальні дані мають певний лінійний зв'язок між випадковими величинами  $x$  та  $y$ , тоді як на рис. 1.1б) кореляційного зв'язку між величинами майже не виявлено (або він є дуже слабким).

Отже, за виглядом кореляційного поля можна передбачити вид зв'язку між досліджуваними величинами та встановити, наприклад, чи є лінійною (або криволінійною) залежність між ними, оцінити величину коефіцієнта кореляції  $r$ . При цьому, чим тісніше точки поля групуються навколо деякої лінії (прямої або кривої), тим ближчою є величина коефіцієнта кореляції до одиниці. І навпаки, чим більше кореляційне поле нагадує коло (тобто неможна виділити домінуючого напрямку), тим ближчою є величина коефіцієнта кореляції до нуля (рис. 1.1б).

За величиною коефіцієнта кореляції  $r$  можна оцінити не тільки тісноту кореляційного зв'язку, а й напрямок кореляції (пряма чи зворотна). Зокрема, при  $r > 0$  кореляційний зв'язок оцінюють як прямий; а при  $r < 0$  – як зворотній.

Таким чином, з урахуванням напрямку коефіцієнт кореляції  $r$  може змінюватись у межах від  $-1$  до  $+1$ . Причому, коли  $r = 1$ , точки лежать на одній прямій. Тоді це означає, що величини  $x$  і  $y$  зв'язані лінійним функціональним зв'язком.

Якщо  $r$  прямує до нуля ( $r \rightarrow 0$ ), то кореляційне поле не має чіткої конфігурації (див. рис.1.1б). Тому можна стверджувати, що між величинами  $x$  та  $y$  не існує лінійного кореляційного зв'язку. Проте у такому випадку між досліджуваними величинами може існувати *нелінійна регресія*.

Оскільки внаслідок статистичного характеру зв'язку одному значенню  $x$  може відповідати сукупність значень  $y$ , то для описання досліджуваного процесу (явища) оптимальною буде функція, для якої виконується умова:

$$S = \sum_{i=1}^n (y_i - \bar{y})^2 \rightarrow \min, \quad (1.2)$$

де  $y_i$  – фактичні ординати кореляційного поля;

$\bar{y}$  – середнє значення ординати з абсцисою  $x$ , розрахованою за апроксимуючим рівнянням.

Отже, регресійний аналіз ґрунтується на *методі найменших квадратів* (МНК). Цей метод, як правило, застосовують для знаходження математичного очікування (сподівання) нормально розподіленої випадкової величини. Якщо на кореляційне поле (див. рис. 1.1а) нанести середні значення  $\bar{y}$  (позначені хрестиками), то лінія  $AB$  буде відповідати функціональній залежності виду:

$$\bar{y} = a + b \cdot x, \quad (\text{або } \bar{y} = a \cdot x + b), \quad (1.3)$$

де  $a$  – коефіцієнт регресії, початок відліку;

$b$  – коефіцієнт регресії, який показує середню зміну залежної змінної при зміні незалежної змінної на одиницю.

Середню лінію кореляційного поля, для якої виконується умова (1.1), називають *лінією регресії*, а числові коефіцієнти, що входять до виразу цього рівняння, – *коефіцієнтами регресії*.

Отже, *рівняння регресії* – це аналітичне рівняння, за допомогою якого визначають взаємозв'язок між досліджуваними ознаками; це регресійна модель залежності результативної ознаки  $y$  від ознаки факторної  $x$ .

При аналізі кореляційної залежності між величинами  $x$  та  $y$ , значення яких визначені для елементів певної вибірки, постають такі завдання:

- вибір виду функції регресії;
- визначення коефіцієнтів регресії.

Найпростішою і найбільш розповсюдженою є лінійна кореляція. У цьому випадку кореляційне поле апроксимують за допомогою рівняння першої степені (лінійне рівняння), а лінію регресії розраховують, використовуючи метод найменших квадратів за умови (1.2).

Рівняння лінійної регресії можна представити або за допомогою рівняння прямої (1.3), або, наприклад, таким чином:

$$y = \bar{y} + |r| \cdot \frac{\delta(y)}{\delta(x)} \cdot (x - \bar{x}) \quad (1.4)$$

де  $\bar{y}$  – середнє значення результативної ознаки (залежної змінної);

$x$  – значення факторної ознаки (незалежної змінної);

$\bar{x}$  – середнє значення факторної ознаки;

$r$  – коефіцієнт парної кореляції.

За умови нормального розподілу величин  $x$  і  $y$  лінії регресії як  $x$  від  $y$ , так і  $y$  від  $x$  є прямими лініями.

Для визначення коефіцієнтів рівняння регресії " $a$ " і " $b$ " знаходять часткові похідні від  $S(a, b)$  по  $a$  і  $b$ , прирівнюють їх до нуля (похідна функції у точці екстремуму дорівнює нулю). Отже, запишемо систему рівнянь:

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (bx_i + a - y_i) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (bx_i + a - y_i)x_i = 0 \end{cases}$$

або

$$\begin{cases} b \sum_{i=1}^n x_i^2 + a \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ b \sum_{i=1}^n x_i + a \cdot n = \sum_{i=1}^n y_i \end{cases}$$

Розв'язуючи систему рівнянь, отримуємо коефіцієнти рівняння регресії "a" і "b" за допомогою виразів (1.5 і 1.6):

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}, \quad (1.5)$$

$$a = y - b \cdot x = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}, \quad (1.6)$$

Рівняння кореляційного зв'язку можна використовувати, наприклад, для обчислення теоретичної лінії регресії, а також очікуваних (теоретичних) чи прогнозованих значень залежної змінної при тих або інших значеннях аргументів.

Залежно від ступеня адекватності кореляційної моделі буде збільшуватися (або зменшуватися) ступінь наближення розрахункових значень результативної ознаки до її фактичного значення. Тому, якщо кореляційна модель враховує всі основні фактори (чинники), що визначають варіацію результативної ознаки, то її точність буде досить високою.

Також при проведенні досліджень виникає необхідність використання й інших кореляційних характеристик. Зокрема:

- *стандартизованого коефіцієнту регресії (бета-коефіцієнту)  $\beta$*  – коефіцієнт регресії, виражений у стандартизованому масштабі і розрахований за формулою (1.7):

$$\beta = b \frac{\sigma_x}{\sigma_y} \quad (1.7)$$

де  $b$  – коефіцієнт чистої регресії при  $i$ -му факторі;

$\sigma_x$  і  $\sigma_y$  – середні квадратичні відхилення, відповідно, по  $i$ -му фактору і результативній ознаці;

- *коефіцієнту еластичності ( $E$ )*, що характеризує ступінь і вагомість впливу конкретного фактора на варіацію результативної ознаки та показує на скільки відсотків у середньому змінюється результативна ознака за умови зміни факторної на 1 %:

$$E = b \frac{\bar{x}}{\bar{y}} \quad (1.8)$$

де  $b$  – коефіцієнт чистої регресії при  $i$ -му факторі;

$\bar{x}$  і  $\bar{y}$  – середні значення, відповідно,  $i$ -го фактора і результативної ознаки.

Для оцінки тісноти або сили зв'язку використовують *коефіцієнт кореляції*. Причому, для різних форм зв'язку застосовують різні показники його тісноти, наприклад:

- при лінійному зв'язку:
  - парний коефіцієнт кореляції  $r$  (при парному зв'язку) та
  - сукупний коефіцієнт кореляції  $R$  (при множинному зв'язку);
- при криволінійному зв'язку:
  - індекс кореляції  $i$  (при парному зв'язку) та
  - кореляційне відношення  $\eta$  (при множинному зв'язку).



Тіснота зв'язку визначається через відношення двох дисперсій:

$$R(r, i, \eta) = \sqrt{\frac{\sigma_0^2}{\sigma_i^2}} \quad (1.9)$$

де  $\sigma_i^2 = \overline{\sum y - (y)^2}$  – загальна дисперсія (загальна варіація результативної ознаки);

$\sigma_0^2$  – дисперсія факторної ознаки.

Як було зазначено, знак коефіцієнту кореляції  $r$  – "+" вказує на прямий, а знак "-" – на зворотній зв'язок між досліджуваними величинами. При цьому важливим є те, що знаки для коефіцієнтів регресії і кореляції повинні збігатися. Якщо ці знаки не збігаються, то це вказує на недостовірні розрахунки, які можуть бути пов'язані з:

- неправильним врахуванням чинників, що впливають на залежну ознаку;
- неправильним аналітичним вираженням зв'язку;
- недостатньою чисельністю вибірки або її неоднорідністю.

Для розрахунку парного коефіцієнта кореляції, як правило, застосовуються формули (1.10) – (1.12):

$$r = \pm \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \left[ n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}}, \quad (1.10)$$

$$r = \pm \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \delta(x) \cdot \delta(y)}, \quad (1.11)$$

$$r = \pm \frac{\left[ \frac{1}{n} \sum_{i=1}^n (x_i y_i) \right] - \bar{x} \cdot \bar{y}}{\delta(x) \cdot \delta(y)}, \quad (1.12)$$

$$\text{де } \delta(x) = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}; \quad \delta(y) = \sqrt{\frac{\sum (y - \bar{y})^2}{n}}.$$

Тісноту зв'язку між випадковими величинами  $x$  та  $y$  вважають *слабкою*, якщо  $|r| < 0,3$ ; тіснота зв'язку є *задовільною*, якщо  $|r| \geq 0,5$ ; якщо  $|r|$  лежить у межах від 0,5 до 0,7 – зв'язок між величинами вважають *досить тісним*; при  $|r| > 0,7$  можна стверджувати, що між досліджуваними ознаками існує *дуже тісний* кореляційний зв'язок.

При проведенні екологічних експериментів (досліджень, спостережень тощо) необхідно враховувати, що на отриманий результат впливають безліч різноманітних факторів. Тому важливо визначати *істотність впливу* кожного з факторів на досліджувану результативну ознаку (величину). Для визначення відсотка мінливості досліджуваної ознаки у щодо її середнього значення, яке визначається мінливістю фактора  $x$ , розраховують *коефіцієнт детермінації*:

$$k_g = r^2 \quad (1.13)$$

де  $r$  – коефіцієнт парної кореляції.

Для того, щоб оцінити тісноту (силу) зв'язку між досліджуваними величинами, коли між ними існує нелінійний зв'язок (або дані вибірки розподілені не за нормальним законом), рекомендується застосовувати *коефіцієнт кореляції Спірмена*. При цьому його статистична значущість перевіряється аналогічно до коефіцієнту кореляції Пірсона.

## 1.2 Оцінка значущості коефіцієнта кореляції

Достовірність (значущість) коефіцієнта кореляції оцінюють за допомогою *критерію Стьюдента*, який розраховують за формулою:

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} \quad (1.14)$$

де  $n$  – кількість експериментальних точок (варіант).

Розраховані значення коефіцієнта Стюдента порівнюють з його граничними (табличними) значеннями  $t_{0,5}; t_{0,1}; t_{0,01}$ , що відповідають рівням значущості  $\alpha = 0,05; 0,01; 0,001$  (додаток А). При цьому кількість ступенів свободи розраховують як  $f = n - 2$ .

Отже, за умови  $t < t_{0,5}$  коефіцієнт кореляції вважають незначимим при рівні значущості  $\alpha = 0,05$  (достовірність  $p \geq 0,95$ ); за умови  $t \geq t_{0,5}$  коефіцієнт кореляції вважають значимим при рівні значущості  $\alpha = 0,05$ . Статистична значущість коефіцієнта кореляції збільшується при  $t > t_{0,1}$  або  $t > t_{0,01}$  (відповідно, з достовірністю  $p \geq 0,99$  або  $p \geq 0,999$ ). Таким чином, під час використання кореляційного методу визначають:

- рівняння та коефіцієнти регресії;
- стандартизований коефіцієнт регресії;
- коефіцієнт еластичності та
- коефіцієнти та індекси кореляції і детермінації.

При цьому одержані рівняння регресії застосовують для:

- дослідження характеру зміни ознаки (величини)  $y$  при зміні фактора (величини)  $x$ ;
- прогнозування значень величини  $y$  при заданих значеннях  $x$ ;
- визначення оптимального значення фактора  $x$  для одержання необхідного значення параметра  $y$  або для утримання параметра  $y$  на певному (зазначеному) рівні.

## РОЗДІЛ 2

### ТЕОРЕТИЧНІ ОСНОВИ РЕГРЕСІЙНОГО АНАЛІЗУ

На стан навколишнього середовища впливають різноманітні, причому часто взаємопов'язані між собою фактори, які діють з різною силою і можуть бути різними за напрямками впливу. Тому на практиці часто виникає потреба у встановленні зв'язку між величиною (ознакою) у та багатьма параметрами  $x_1, x_2, \dots, x_n$ .

Як зазначалось, у більшості випадків між змінними існують залежності, при яких кожному значенню однієї величини (аргументу) відповідає безліч можливих значень іншої величини (тобто її розподіл). Така залежність називається *стохастичною* або *ймовірнісною*. При цьому окремим випадком стохастичної залежності вважається кореляційна.

Наявність кореляційної залежності між змінними не завжди означає існування між ними *безпосереднього* зв'язку. Зокрема, зв'язок може існувати і завдяки іншим чинникам, які не було враховано. Крім того, досліджувані величини можуть бути пов'язані через латентні, так звані приховані від дослідника змінні. Таке явище часто спостерігається, наприклад, у педагогічних, психологічних та соціологічних дослідженнях. Можливим також є встановлення випадкової кореляції, яка не обумовлена жодною з причин, виокремлених дослідником.

Отже, кореляційний зв'язок показує тенденцію зміни однієї величини під дією іншої. Тому на підставі кореляції можна стверджувати тільки про *ступінь зв'язку* між досліджуваними змінними, а не про існування між ними причинно-наслідкового зв'язку.

Рівняння зв'язку може бути знайдене не тільки для парних залежностей, а й для множинного зв'язку. Такі завдання зазвичай

розв'язують, використовуючи методи *багатофакторної регресії* (багатофакторний кореляційно-регресійний аналіз).

Отже, багатофакторний кореляційний-регресійний аналіз застосовують для таких цілей:

- розрахунку теоретичних (очікуваних) значень результативної ознаки;
- зіставлення й оцінювання фактичного та розрахункового значень результативної ознаки;
- порівняльного аналізу різних вибірок;
- розроблення нормативів і прогнозування розвитку екологічних (й інших) систем та явищ.

Основною відмінністю між парною та багатофакторною кореляціями є те, що остання дає змогу оцінити ступінь впливу на результативну ознаку кожного з досліджуваних факторів при фіксованих значеннях решти чинників. Окрім того, як показує практика, парна кореляція не завжди дає правильне уявлення про зв'язок між результативною і факторною ознаками.

Важливою умовою отримання достовірних результатів щодо наявності й виду зв'язку між досліджуваними ознаками є вибір факторів, які найістотніше впливають на результативну ознаку. Це обумовлено тим, що одночасне врахування великої кількості різноманітних факторів (особливо при відносно невеликій чисельності вибірки) може призвести до помилкових результатів та висновків.

Окрім того, при збільшенні в рівнянні регресії кількості досліджуваних параметрів значно ускладнюється тлумачення одержаних результатів. Ось чому потрібно оцінити й відібрати найбільш значущі фактори, які чинять найсильніший вплив на результативну ознаку. Для зменшення кількості параметрів рекомендують застосовувати, наприклад, метод статистичного групування.

Головною проблемою при побудові рівняння множинної регресії є вибір адекватного математичного рівняння (функції), яке б повністю відповідало характеру взаємозв'язку між результативною ознакою і включеними до рівняння регресії змінними. При цьому прийоми вибору виду зв'язку, які використовуються при парній кореляції, як правило, не можуть застосовуватися для множинної регресії. Тому для оцінки тісноти зв'язку між конкретною ознакою  $y$  й усіма іншими ознаками  $(x_1, x_2, \dots, x_n)$  застосовують *множинний коефіцієнт кореляції*, який позначається  $R$ . Для його розрахунку потрібно побудувати матрицю парних коефіцієнтів кореляції  $r_{ij}$  між досліджуваними ознаками, знайти визначник  $|A|$  матриці  $A$  та алгебраїчне доповнення  $A_{ii}$  елемента  $r_{ii}$  цієї матриці. Далі множинний коефіцієнт кореляції розраховують за формулою:

$$R_i = (1 - |A|/A_{ii})^{1/2} \quad . \quad (2.1)$$

Перевірку статистичної значущості множинного коефіцієнта  $R$  здійснюють за допомогою  $F$ -статистики шляхом порівняння розрахованого значення з критичним значенням  $F_{\text{крит}}$ . ( $F_{\text{крит}}$  – табличне значення розподілу Фішера, додаток Б). Якщо розраховане значення  $F$ -статистики за абсолютною величиною більше за критичне значення, то множинний коефіцієнт кореляції вважають значущим на обраному рівні значущості  $\alpha$ .

У деяких випадках необхідно оперувати *частинними коефіцієнтами кореляції*  $R_{ij}$ . Для розрахунку цього коефіцієнту також потрібно побудувати матрицю парних коефіцієнтів кореляції  $A$  та знайти алгебраїчні доповнення  $A_{ii}$ ,  $A_{jj}$ ,  $A_{ij}$  елементів  $r_{ii}$ ,  $r_{jj}$ ,  $r_{ij}$ , відповідно.

Отже, регресійний аналіз проводять за таким алгоритмом:

- встановлюють вид кореляційної залежності результативної ознаки  $y$  від факторної ознаки  $x$ ;
- будують регресійну модель та
- перевіряють статистичну значущість побудованої моделі.

Для вибору рівняння регресії застосовують такі методи, як побудова комбінаційних групувань, експертні оцінки, вивчення парних зв'язків між результативною ознакою і кожним з досліджуваних факторів, побудова графіків, перебирання функцій різних типів та ін. Проте всі ці методи, як правило, призводять до великої кількості додаткових обчислень. При цьому у більшості випадків кореляційні зв'язки досить добре описуються функціями лінійного типу або степеневими функціями, які логарифмуванням або методом заміни змінних можна привести до лінійного вигляду.

У загальному випадку лінійна багатофакторна регресія може бути записана як:

$$y = f(x_1, x_2, \dots, x_n) + \varepsilon, \quad (2.2)$$

де  $y$  – результативна ознака (залежна змінна);

$x_1, x_2, \dots, x_n$  – факторні ознаки (незалежні змінні);

$n$  – кількість факторних ознак;

$\varepsilon$  – випадкова похибка моделі.

Таким чином, при  $n$  змінних лінійне рівняння множинної регресії має вигляд:

$$\tilde{y}_x = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n, \quad (2.3)$$

де  $\tilde{y}_x$  – залежна змінна (результативна ознака);

$x_i$  – незалежні змінні (фактори);

$a_0$  – коефіцієнт регресії, початок відліку;

$a_1, a_2, \dots, a_n$  – інші коефіцієнти регресії.

Таким чином, *рівняння множинної регресії* – це математичне рівняння (модель), яким виражається кореляційний зв'язок між декількома досліджуваними ознаками. Параметри рівняння регресії також знаходять використовуючи метод найменших квадратів.

Зазначимо, що для побудови адекватної багатofакторної регресійної моделі необхідно, щоб випадкова похибка  $\varepsilon$  мала нормальний розподіл із нульовим математичним очікуванням, а випадкові похибки кожного вимірювання були незалежними і мали однакові дисперсії. Окрім того, кількість спостережень  $N$  повинна перевищувати величину  $3(n + 1)$ .

Найважливішою умовою побудови коректних багатofакторних регресійних моделей і забезпечення їх статистичної значущості є те, що всі досліджувані факторні ознаки повинні бути тісно пов'язані з результативною ознакою, проте бути незалежними одна від одної. Тісноту зв'язку між результативною і факторними ознаками, а також зв'язку факторних ознак між собою визначають аналізом парних і частинних коефіцієнтів кореляції.

Коефіцієнти множинної регресії показують ступінь середньої зміни результативної ознаки за умови зміни відповідної факторної ознаки на одиницю коли всі інші фактори (або ознаки), які включені до рівняння регресії, залишаються сталими на одному середньому рівні. Отже, *коефіцієнти чистої регресії* – це коефіцієнти множинної регресії, які характеризують зв'язок між результативною ознакою і певним фактором при фіксованих значеннях інших факторів. *Коефіцієнтами повної регресії* називають коефіцієнти парної регресії.



### РОЗДІЛ 3

## АЛГОРИТМИ ТА ПРИКЛАДИ ВИКОНАННЯ

### РОЗРАХУНКІВ (ЛІНІЙНА РЕГРЕСІЯ)

*Приклад № 1.* У результаті проведення експериментального дослідження отримано показники  $x$  та  $y$  (табл. 3.1). Позначимо незалежну ознаку як  $x$ , а залежну – як  $y$ . Проаналізуємо вихідні дані за таким алгоритмом:

- для побудови кореляційного поля, наприклад, у Microsoft Office Excel до табл. 3.1 (стовбці 2 і 3) вносимо значення змінних  $x$  і  $y$ ;
- проводимо розрахунки за формулами, наведеними у табл. 3.1;
- підбираємо рівняння лінійної регресії, розраховуємо коефіцієнти регресії та кореляції, перевіряємо значущість за допомогою критерію Стюдента та аналізуємо тісноту кореляційного зв'язку між досліджуваними ознаками;
- будуємо графік залежності між ознаками  $y$  та  $x$  та порівнюємо його з експериментальними даними.

Отже, вводимо до табл. 3.1 експериментальні значення  $x$  та  $y$  ( $N = 50$ ) і будуємо кореляційне поле (рис. 3.1). Із розташування точок видно, що вони формують певну залежність, яка вірогідно має лінійний характер.

У результаті проведення розрахунків отримуємо такі коефіцієнти регресії  $a$  і  $b$  та коефіцієнт кореляції  $r$ :

$$b = 0,1039$$

$$a = -7,5777$$

$$r = 0,99$$

Оскільки  $r$  наближається до 1, доходимо висновку, що існує дуже тісний кореляційний зв'язок між незалежною ознакою  $x$  та залежною ознакою  $y$ . Отримане рівняння регресії має вигляд:

$$y = -7,5777 + 0,1039 * x$$

Таблиця 3.1 – Експериментальні дані (стовбці 2, 3) і результати розрахунків (стовбці 4–11)

№	$x$	$y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})$	$x^2$	$y^2$	$xy$	$(x - \bar{x})(y - \bar{y})$
<i>l</i>	2	3	4	5	6	7	8	9	10	11
1	131,72	5,48	-5,12	-1,16	26,17	1,34	17350,16	30,03	721,83	5,93
2	126,08	5,32	10,76	-1,32	115,68	1,74	15896,17	28,30	670,75	14,19
3	142,95	6,85	6,11	0,21	37,39	0,04	20434,70	46,92	979,21	1,29
4	103,07	3,46	33,77	-3,18	1140,10	10,11	10623,42	11,97	356,62	107,35
5	101,47	3,41	35,37	-3,23	1250,71	10,43	10296,16	11,63	346,01	114,21
6	125,03	5,04	11,81	-1,60	139,37	2,56	15632,50	25,40	630,15	18,88
7	126,93	5,41	-9,91	-1,23	98,12	1,51	16111,22	29,27	686,69	12,18
8	123,81	4,79	13,03	-1,85	169,66	3,42	15328,92	22,94	593,05	24,09
9	141,14	6,91	4,30	0,27	18,53	0,07	19920,50	47,75	975,28	1,16
10	150,86	7,82	14,02	1,18	196,69	1,39	22758,74	61,15	1179,73	16,56
11	129,62	5,86	-7,22	-0,78	52,06	0,61	16801,34	34,34	759,57	5,62
12	133,66	6,08	-3,18	-0,56	10,08	0,31	17865,00	36,97	812,65	1,78
13	132,81	6,01	-4,03	-0,63	16,20	0,40	17638,50	36,12	798,19	2,53
14	122,26	4,86	14,58	-1,78	212,44	3,17	14947,51	23,62	594,18	25,94
15	107,92	3,68	28,92	-2,96	836,10	8,76	11646,73	13,54	397,15	85,57
16	103,36	3,39	33,48	-3,25	1120,60	10,56	10683,29	11,49	350,39	108,77
17	137,54	6,21	0,70	-0,43	0,50	0,18	18917,25	38,56	854,12	-0,30
18	121,18	4,97	15,66	-1,67	245,09	2,79	14684,59	24,70	602,26	26,14
19	145,79	7,61	8,95	0,97	80,18	0,94	21254,72	57,91	1109,46	8,69
20	110,52	4,15	26,32	-2,49	692,50	6,20	12214,67	17,22	458,66	65,51
21	160,03	8,72	23,19	2,08	537,99	4,33	25609,60	76,04	1395,46	48,26
22	159,08	8,51	22,24	1,87	494,82	3,50	25306,45	72,42	1353,77	41,61
23	130,05	5,71	-6,79	-0,93	46,04	0,86	16913,00	32,60	742,59	6,31
24	128,06	5,82	-8,78	-0,82	77,01	0,67	16399,36	33,87	745,31	7,19
25	126,94	5,03	-9,90	-1,61	97,92	2,59	16113,76	25,30	638,51	15,93
26	111,96	4,11	24,88	-2,53	618,79	6,40	12535,04	16,89	460,16	62,92
27	143,29	7,49	6,45	0,85	41,66	0,72	20532,02	56,10	1073,24	5,49
28	158,46	9,04	21,62	2,40	467,62	5,76	25109,57	81,72	1432,48	51,91
29	161,03	9,26	24,19	2,62	585,38	6,87	25930,66	85,75	1491,14	63,40

№	$x$	$y$	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})$	$x^2$	$y^2$	$xy$	$(x - \bar{x})(y - \bar{y})$
1	2	3	4	5	6	7	8	9	10	11
30	165,39	9,91	28,55	3,27	815,37	10,70	27353,85	98,21	1639,01	93,39
31	146,98	8,38	10,14	1,74	102,91	3,03	21603,12	70,22	1231,69	17,66
32	143,27	6,97	6,43	0,33	41,40	0,11	20526,29	48,58	998,59	2,13
33	170,72	10,31	33,88	3,67	1148,17	13,47	29145,32	106,30	1760,12	124,38
34	168,93	10,24	32,09	3,60	1030,06	12,96	28537,34	104,86	1729,84	115,56
35	150,36	7,48	13,52	0,84	182,91	0,71	22608,13	55,95	1124,69	11,37
36	124,93	4,91	11,91	-1,73	141,74	2,99	15607,50	24,11	613,41	20,59
37	167,78	9,89	30,94	3,25	957,57	10,57	28150,13	97,81	1659,34	100,59
38	163,66	9,57	26,82	2,93	719,56	8,59	26784,60	91,58	1566,23	78,61
39	134,29	6,02	-2,55	-0,62	6,48	0,38	18033,80	36,24	808,43	1,58
40	171,03	10,53	34,19	3,89	1169,27	15,14	29251,26	110,88	1800,95	133,04
41	168,96	10,27	32,12	3,63	1031,99	13,18	28547,48	105,47	1735,22	116,63
42	125,09	4,99	11,75	-1,65	137,95	2,72	15647,51	24,90	624,20	19,37
43	98,02	3,07	38,82	-3,57	1506,64	12,74	9607,92	9,42	300,92	138,55
44	97,36	3,09	39,48	-3,55	1558,31	12,60	9478,97	9,55	300,84	140,11
45	101,73	3,31	35,11	-3,33	1232,39	11,08	10348,99	10,96	336,73	116,88
46	163,74	10,11	26,90	3,47	723,86	12,05	26810,79	102,21	1655,41	93,38
47	99,53	3,27	37,31	-3,37	1391,69	11,35	9906,22	10,69	325,46	125,70
48	149,59	8,57	12,75	1,93	162,68	3,73	22377,17	73,44	1281,99	24,62
49	160,48	9,35	23,64	2,71	559,07	7,35	25753,83	87,42	1500,49	64,09
50	173,31	10,73	36,47	4,09	1330,40	16,73	30036,36	115,13	1859,62	149,20
-	Сума	Сума			Сума	Сума	Сума	Сума	Сума	Сума
	6841,77	331,97			25375,82	280,42	961572,16	2484,50	48061,78	2636,53
	$\bar{x}$	$\bar{y}$								
	136,84	6,64								

Для перевірки розрахунків розв'язуємо обернену задачу – результат розрахунку оберненої задачі наведено у табл. 3.2. Як можна побачити, значних розбіжностей між експериментальними і розрахунковими даними не спостерігається.

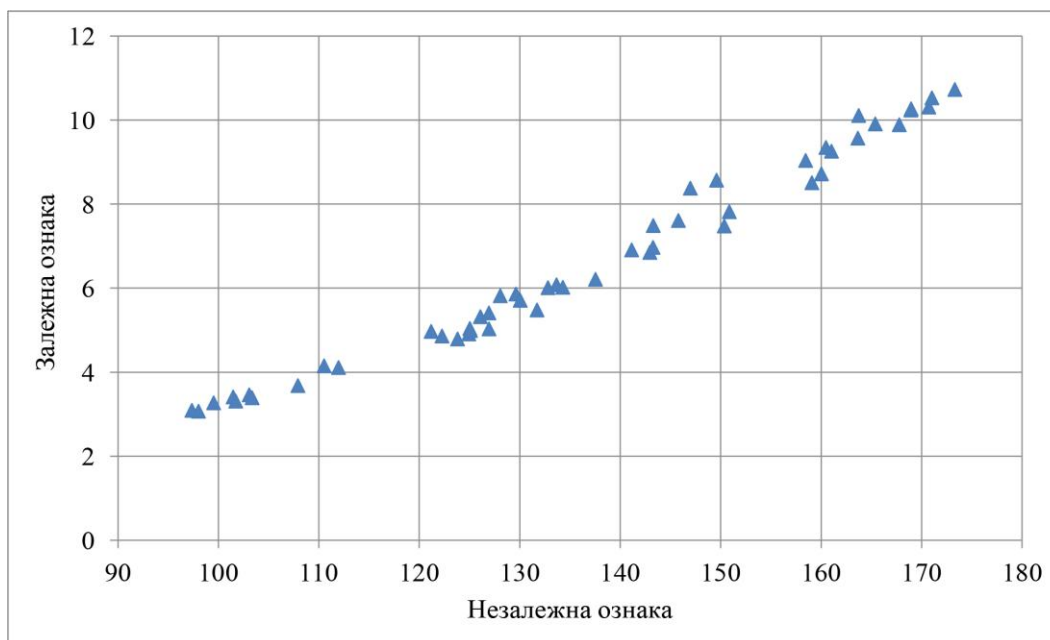


Рисунок 3.1 – Вид кореляційного поля для прикладу № 1.

Таблиця 3.2 – Результат перевірки правильності розрахунків

№	$y$	$y'$	№	$y$	$y'$
1.	5,48	6,11	26.	4,11	4,05
2.	5,32	5,52	27.	7,49	7,31
3.	6,85	7,27	28.	9,04	8,89
4.	3,46	3,13	29.	9,26	9,15
5.	3,41	2,96	30.	9,91	9,61
6.	5,04	5,41	31.	8,38	7,69
7.	5,41	5,61	32.	6,97	7,31
8.	4,79	5,29	33.	10,31	10,16
9.	6,91	7,09	34.	10,24	9,97
10.	7,82	8,10	35.	7,48	8,04
11.	5,86	5,89	36.	4,91	5,40
12.	6,08	6,31	37.	9,89	9,85
13.	6,01	6,22	38.	9,57	9,43
14.	4,86	5,13	39.	6,02	6,37
15.	3,68	3,64	40.	10,53	10,19
16.	3,39	3,16	41.	10,27	9,98
17.	6,21	6,71	42.	4,99	5,42
18.	4,97	5,01	43.	3,07	2,61
19.	7,61	7,57	44.	3,09	2,54
20.	4,15	3,91	45.	3,31	2,99
21.	8,72	9,05	46.	10,11	9,43
22.	8,51	8,95	47.	3,27	2,76
23.	5,71	5,93	48.	8,57	7,96
24.	5,82	5,73	49.	9,35	9,10
25.	5,03	5,61	50.	10,73	10,43

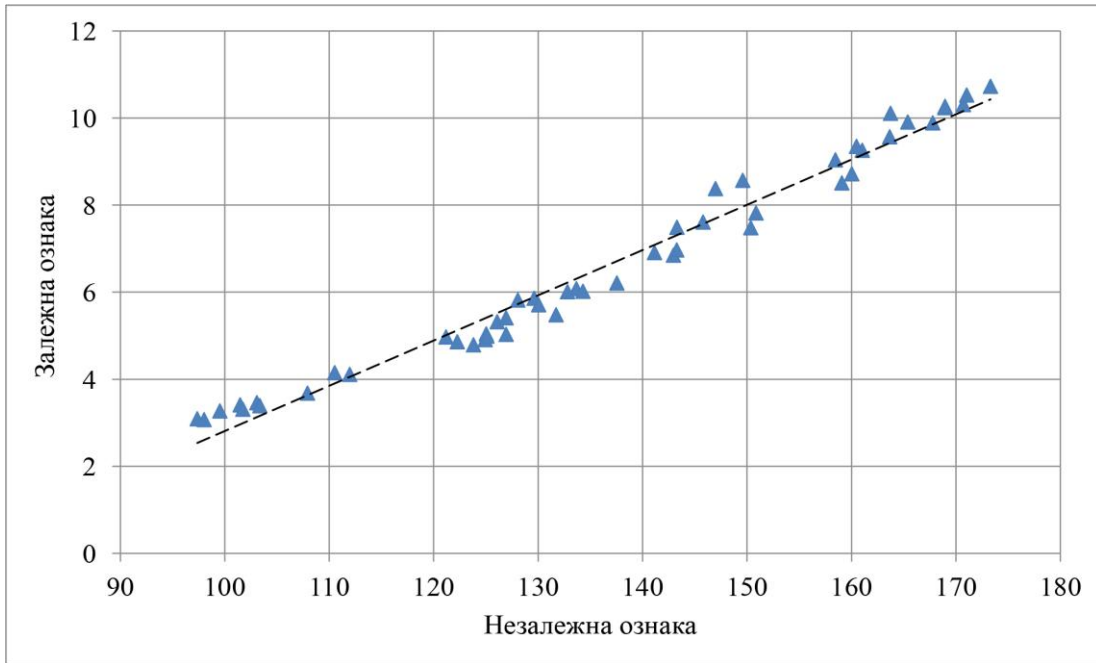


Рисунок 3.2. Залежність між ознаками  $y$  та  $x$  для прикладу № 1.

*Приклад № 2.* У дослідженні отримано дані для  $x$  та  $y$  (рис. 3.3). Аналіз цих даних виконуємо аналогічно до попереднього прикладу.

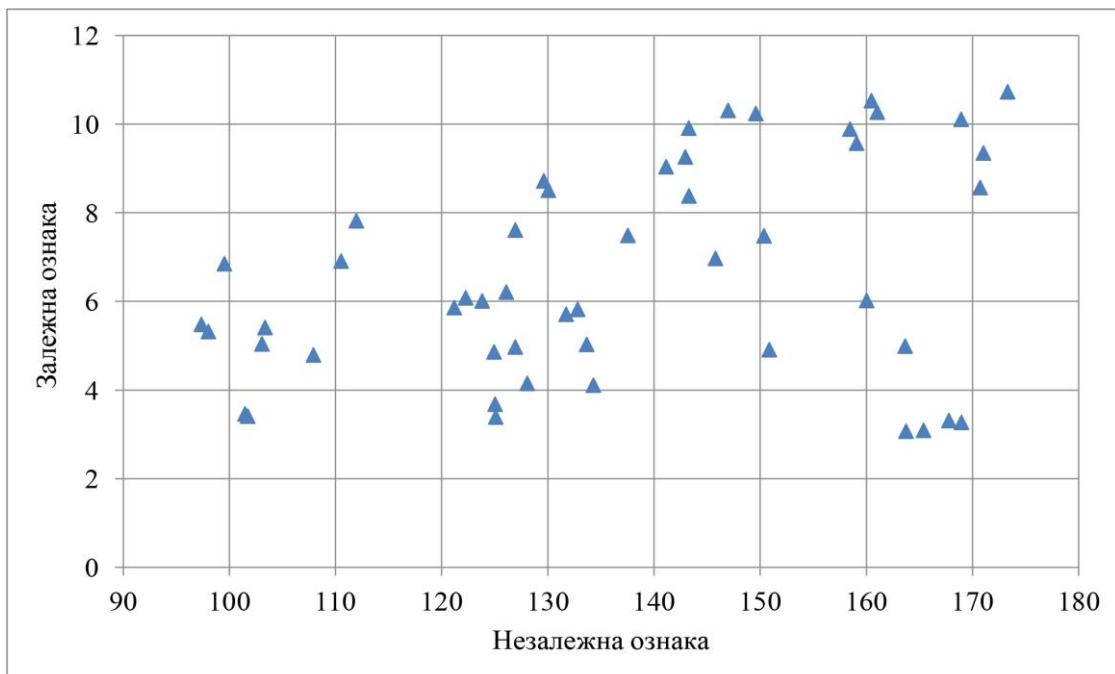


Рисунок 3.3 – Вид кореляційного поля для прикладу № 2.

Із розташування точок видно, що існує певна залежність між досліджуваними ознаками, однак чіткої лінійної залежності не помітно. Визначаємо шукані коефіцієнти рівняння регресії та коефіцієнт кореляції, а також перевіряємо значущість коефіцієнта кореляції за допомогою критерію Стьюдента:

$$b = 0,0384$$

$$a = 1,3849$$

$$r = 0,37$$

За величиною коефіцієнта кореляції  $r$  доходимо висновку, що зв'язок між ознаками помірний. Рівняння регресії має вигляд:

$$y = 1,3849 + 0,0384*x$$

У результаті розв'язання оберненої задачі встановлюємо, що значних розбіжностей між експериментальними та розрахунковими значеннями ( $y$  та  $y'$ ) немає (рис. 3.4). Отже, отримані результати свідчать про коректність розрахунків.

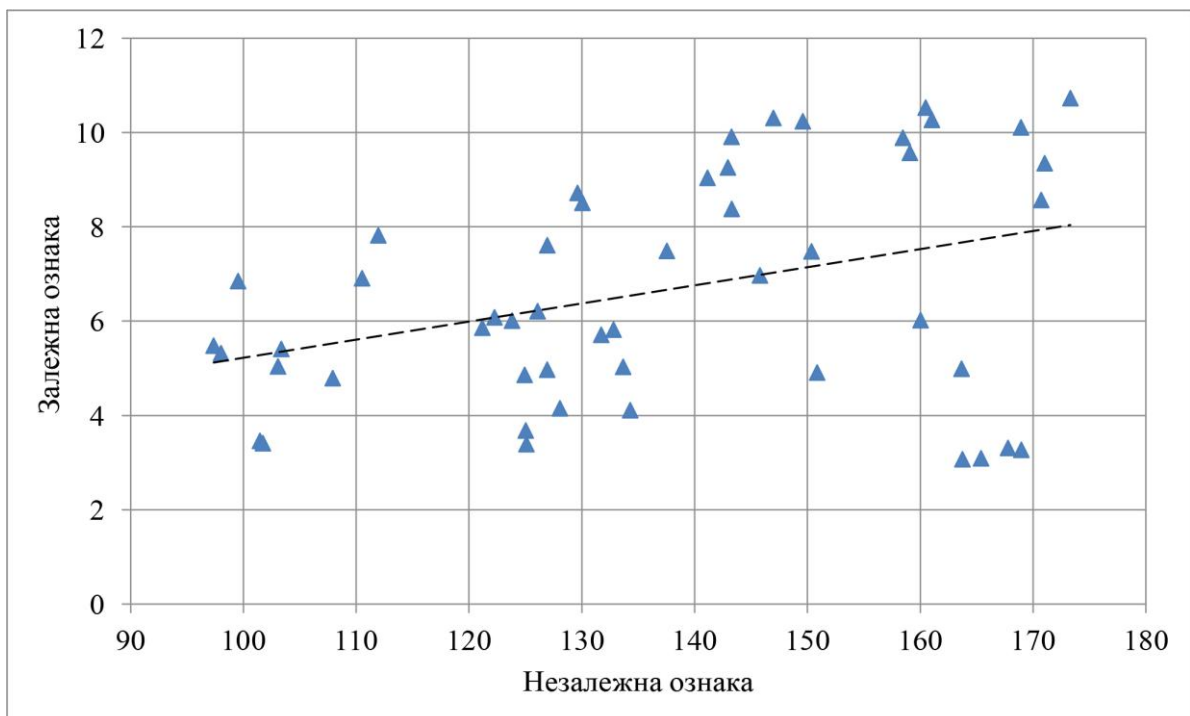


Рисунок 3.4 – Залежність ознаки  $y$  від ознаки  $x$  для прикладу № 2.

Приклад № 3. Шляхом проведення експериментального дослідження отримано дані для  $y$  та  $x$ . Аналіз і статистичну обробку результатів дослідження виконуємо аналогічно до попередніх прикладів.

На рис. 3.5 показано побудоване за експериментальними даними кореляційне поле. Із розташування точок видно, що залежності між величинами не простежується.

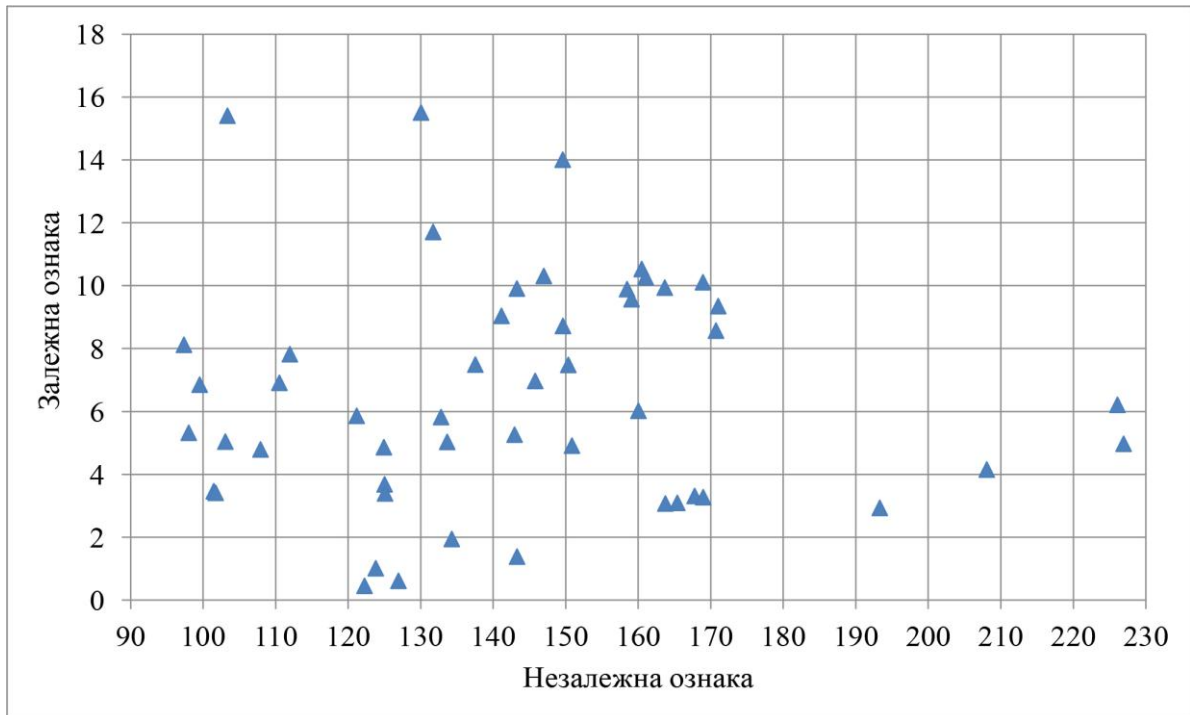


Рисунок 3.5 – Кореляційне поле, побудовано за даними прикладу № 3.

Розрахунок показує, що величина коефіцієнту кореляції становить  $r = 0,02$ . Отже, доходимо висновку, що кореляційний зв'язок між досліджуваними ознаками практично відсутній і у такому випадку визначати коефіцієнти рівняння регресії не потрібно. Проте у подібних випадках існує вірогідність того, що було невірно обрано факторну ознаку або на результативну ознаку у одночасно чинять вплив декілька різних факторів.

## РОЗДІЛ 4

### ПРИКЛАДИ ЗАВДАНЬ ДО САМОСТІЙНОЇ РОБОТИ

*Завдання № 1.* У результаті проведеного дослідження отримано показники  $x$  та  $y$  (табл. 4.1). Відповідно до наданого викладачем варіанту провести розрахунки і проаналізувати результати дослідження.

Таблиця 4.1 – Експериментальні дані для виконання завдання № 1

№	Варіант № 1		Варіант № 2		Варіант № 3		Варіант № 4		Варіант № 5	
	$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$
1	157,92	8,07	109,01	3,98	100,91	3,02	104,52	3,58	107,84	4,11
2	160,03	8,72	150,68	8,14	125,47	5,18	163,74	10,11	149,03	7,15
3	159,08	8,51	130,02	5,79	129,62	5,86	99,53	3,27	103,36	3,39
4	123,81	4,79	129,62	5,86	133,66	6,08	149,59	8,57	137,54	6,21
5	141,14	6,91	133,66	6,08	132,81	6,01	160,48	9,35	121,18	4,97
6	150,86	7,82	132,81	6,01	122,26	4,86	173,31	10,73	145,79	7,61
7	129,62	5,86	122,26	4,86	107,92	3,68	133,66	6,08	110,52	4,15
8	122,26	4,86	107,92	3,68	103,36	3,39	132,81	6,01	160,03	8,72
9	107,92	3,68	103,36	3,39	137,54	6,21	122,26	4,86	168,93	10,24
10	103,36	3,39	137,54	6,21	121,18	4,97	107,92	3,68	150,36	7,48
11	137,54	6,21	121,18	4,97	145,79	7,61	103,36	3,39	124,93	4,91
12	134,29	6,02	145,79	7,61	110,52	4,15	137,54	6,21	97,36	3,09
13	143,27	6,97	110,52	4,15	160,03	8,72	121,18	4,97	101,73	3,31
14	170,72	10,31	165,39	9,91	159,08	8,51	145,79	7,61	163,74	10,11
15	168,93	10,24	146,98	8,38	130,05	5,71	128,06	5,82	134,29	6,02
16	168,96	10,27	143,27	6,97	161,03	9,26	126,94	5,03	171,03	10,53
17	125,09	4,99	170,72	10,31	165,39	9,91	111,96	4,11	168,96	10,27
18	98,02	3,07	168,93	10,24	146,98	8,38	143,29	7,49	131,72	5,48
19	97,36	3,09	150,36	7,48	143,27	6,97	158,46	9,04	126,08	5,32
20	150,36	7,48	124,93	4,91	170,72	10,31	161,03	9,26	142,95	6,85
21	124,93	4,91	98,02	3,07	163,66	9,57	165,39	9,91	103,07	3,46
22	143,29	7,49	97,36	3,09	134,29	6,02	146,98	8,38	101,47	3,41
23	158,46	9,04	101,73	3,31	99,53	3,27	143,27	6,97	125,03	5,04
24	95,73	2,89	141,81	7,12	97,14	3,51	141,02	6,59	126,93	5,41
25	145,79	7,61	128,06	5,82	107,92	3,68	160,03	8,72	137,54	6,21



№	Варіант № 1		Варіант № 2		Варіант № 3		Варіант № 4		Варіант № 5	
	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
26	110,52	4,15	126,94	5,03	103,36	3,39	159,08	8,51	134,29	6,02
27	160,03	8,72	111,96	4,11	137,54	6,21	130,05	5,71	143,27	6,97
28	168,93	10,24	143,29	7,49	121,18	4,97	161,03	9,26	170,72	10,31
29	126,08	5,32	158,46	9,04	145,79	7,61	165,39	9,91	168,93	10,24
30	129,06	5,33	111,92	4,35	165,07	9,83	163,83	9,57	129,04	6,05

*Завдання № 2.* Використовуючи безкоштовні репозиторії, знайти екологічні дані та виконати розрахунки і аналіз результатів досліджень, користуючись алгоритмом, наведеним у розділі 3.

Репозиторії являють собою бази даних, які містять великі обсяги інформації (структурованої або ні) і які в залежності від умов ліцензії можуть бути використані для проведення наукових досліджень.

Репозиторії у сфері екології та охорони навколишнього середовища, біології, хімії, кліматології тощо є, наприклад, такі:

- <https://knb.ecoinformatics.org/> (рис. 4.1);
- <https://www.ncbi.nlm.nih.gov/taxonomy> (рис. 4.2);
- <https://www.gbif.org/> (рис. 4.3);
- <https://www.ncei.noaa.gov/> (рис. 4.4);
- <https://daac.ornl.gov/> ;
- <http://www.earthchem.org/> ;
- <https://www.pangaea.de/> .

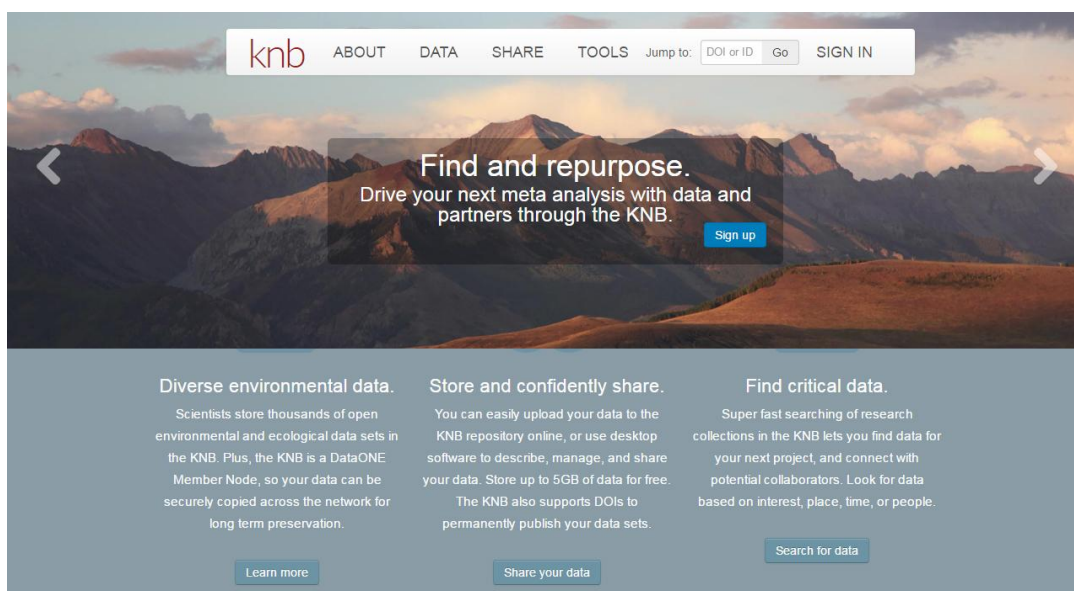


Рисунок 4.1 – Скріншот репозиторію "KNB Ecoinformatics".

The screenshot shows the NCBI Taxonomy homepage. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' menus, and a 'Sign in to NCBI' link. Below this is a search bar with 'Taxonomy' selected in a dropdown menu and a 'Search' button. The main content area features a header with the word 'Taxonomy' and a description: 'The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.' To the left of this text is a grid of colorful butterfly images. Below the header, there are three columns of links: 'Using Taxonomy' (Quick Start Guide, FAQ, Handbook, Taxonomy FTP), 'Taxonomy Tools' (Browser, Common Tree, Statistics, Name/ID Status, Genetic Codes, Linking to Taxonomy, Extinct Organisms), and 'Other Resources' (GenBank, LinkOut, E-Utilities, Batch Entrez, INSDC).

Рисунок 4.2 – Скріншот репозиторію NCBI.

The screenshot displays the GBIF search results page for 'SEARCH DATASETS | 40,668 RESULTS'. The page is divided into a left sidebar and a main content area. The sidebar contains filters for 'Publisher' and 'Host', each with a search bar and a list of options. The main content area shows a list of datasets under the 'ALL' tab, with sub-tabs for 'OCCURRENCE', 'CHECKLIST', 'SAMPLING EVENT', and 'METADATA'. The datasets listed are: 'EOD - eBird Observation Dataset' (361 429 888 occurrences, 26 citations), 'DOF' (19 287 097 occurrences, 11 citations), 'Finnish Bird Ringing and Recovery Database' (10 930 328 occurrences, 1 citation), 'BirdLife Australia, Birdata' (10 153 635 occurrences, 8 citations), and 'Flora von Deutschland (Phanerogamen)'. Each dataset entry includes a brief description, the publisher, and a small map showing the distribution of the data.

Рисунок 4.3 – Скріншот репозиторію GBIF.

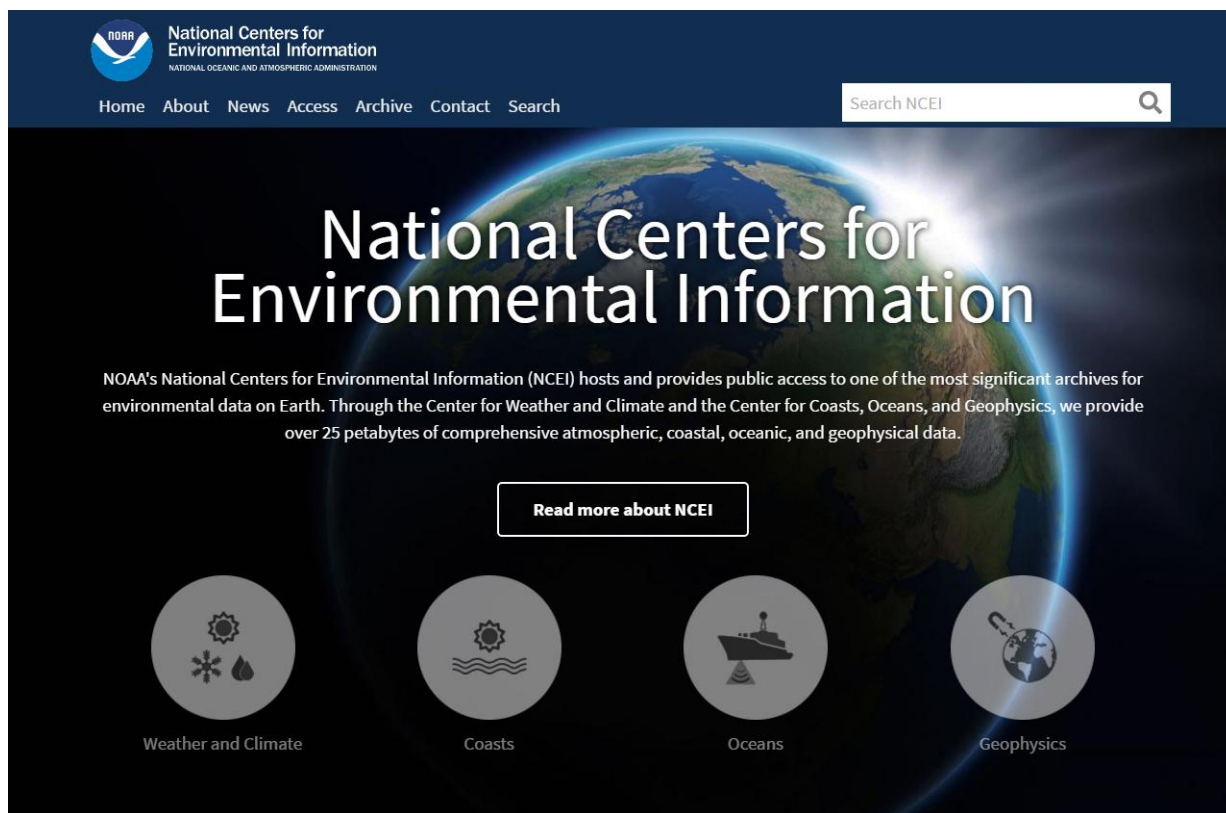


Рисунок 4.4 – Скріншот репозиторію NCEI (National Centers for Environmental Information).

*Завдання № 3.* 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 розрахованих за рівнянням регресії величин  $y$  та порівняти їх з експериментальними.

Варіант № 1. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	6	7	5	6	5	5	4	5	5	4	6
$y$	4	5	4	4	6	2	3	3	2	6	6

Варіант № 2. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	7	6	5	6	6	7	5	6	6	7	3
$y$	9	7	2	4	7	5	6	8	10	4	5

Варіант № 3. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	6	7	4	5	5	6	5	6	4	5	7
$y$	5	4	2	5	6	5	9	3	6	4	3

Варіант № 4. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	7	6	6	6	5	7	6	5	4	6	5
$y$	4	2	5	7	3	5	4	5	3	7	2

Варіант № 5. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	5	6	6	5	6	8	5	6	6	5	6
$y$	3	2	5	3	7	4	2	4	5	3	6

Варіант № 6. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	5	6	7	5	8	8	5	4	6	7	6
$y$	3	2	6	3	7	8	2	4	3	3	7

Варіант № 7. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	6	7	6	7	5	4	5	5	6	5	3
$y$	4	9	5	4	3	3	6	9	4	2	5

Варіант № 8. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	6	6	5	6	5	4	5	5	6	5	7
$y$	4	4	5	7	3	3	6	9	4	2	4

Варіант № 9. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	4	5	5	6	5	5	6	7	6	7	5
$y$	3	6	9	4	2	4	4	9	5	4	3

Варіант № 10. Отримано такі дані щодо плодючості самок лисиць  $x$  та плодючості їх дочок  $y$ :

$x$	5	6	4	4	6	4	7	5	6	5	3
$y$	2	10	6	3	5	6	4	4	7	3	6

## ВИКОРИСТАНА ЛІТЕРАТУРА

1. Василенко О. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. / О. А. Василенко, І. А. Сенча. – Одеса: ОНАЗ ім. О. С. Попова, 2011. – 166 с.
2. Грицевич В. С. Кореляційний та регресійний аналіз в суспільній географії: тексти лекцій / В. С. Грицевич. – Львів: Малий видавничий центр. Лабораторія тематичного картографування географічного факультету, 2016. – 24 с.
3. Грицевич В. С. Статистичні ознаки та характеристики їхньої центральної тенденції: тексти лекцій / В. С. Грицевич. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2008. – 52 с.
4. Захожай В. Б. Практикум з основ статистики / Захожай В. Б. Попов І. І., Коваленко О. В. – К.: МАУП, 2001. – 172 с.
5. Кофанова О. В. Основи статистичного обліку в екології. Модуль 1: Елементи теорії імовірностей та математичної статистики [Електронний ресурс] / О. В. Кофанова, Т. М. Перельот. – 2010. – Режим доступу до ресурсу: <http://ela.kpi.ua/handle/123456789/1850>.
6. Лапач С. Н. Статистика в науке и бизнесе / С. Н. Лапач, А. В. Чубенко, П. Н. Бабич. – К.: МОРИОН, 2002. – 640 с.
7. Матковський С. О. Теорія статистики: навч. посіб. / Матковський С. О., Марець О. Р. – К.: Знання, 2009. – 534 с.
8. Минько А. А. Статистический анализ в MS Excel / Минько А. А. – М.: Изд. дом "Вильямс", 2004. – 448 с.
9. Тарасова В. В. Екологічна статистика (з блочно-модульною формою контролю знань): підручник / Тарасова В. В. – К.: Центр учбової літератури, 2008. – 392 с.
10. Турчин В. М. Теорія ймовірностей і математична статистика. Основні поняття, приклади, задачі : [підруч. для студентів] / В. М. Турчин. – Дніпропетровськ : ІМА-прес, 2014. – 555 с.

11. Харченко М. А. Корреляционный анализ: учеб. пособ. для вузов / М. А. Харченко. – Воронеж: Издательско-полиграфический центр Воронежского гос. ун-та, 2008. – 30 с.
12. Швець В. Т. Теорія ймовірностей і математична статистика : [навч. посіб.] / В. Т. Швець. – Одеса : ВМВ, 2014. – 200 с.
13. Data Publisher for Earth & Environmental Science – Pangaea [Електронний ресурс] : [Сайт]. – Режим доступу : <https://www.pangaea.de>. – Назва з екрана.
14. EarthChem [Електронний ресурс] : [Сайт]. – Режим доступу : <http://www.earthchem.org>. – Назва з екрана.
15. Global Biodiversity Information Facility [Електронний ресурс] : [Сайт]. – Режим доступу : <https://www.gbif.org>. – Назва з екрана.
16. National Centers for Environmental Information (NCEI) [Електронний ресурс] : [Сайт]. – Режим доступу : <https://www.ncei.noaa.gov>. – Назва з екрана.
17. NCBI Taxonomy [Електронний ресурс] : [Сайт]. – Режим доступу : <https://www.ncbi.nlm.nih.gov/taxonomy>. – Назва з екрана.
18. Ofungwu J. Statistical Applications for Environmental Analysis and Risk Assessment (Statistics in Practice) / Joseph Ofungwu. – Hackettstown: John Wiley & Sons, Inc., 2014. – 648 с. – (1).
19. ORNL DAAC for Biogeochemical Dynamics [Електронний ресурс] : [Сайт]. – Режим доступу : <https://daac.ornl.gov>. – Назва з екрана.
20. Schuenemeyer J. H. Statistics for Earth and Environmental Scientists / J. H. Schuenemeyer, L. J. Drew. – Hoboken: John Wiley & Sons, Inc., 2011. – 384 с. – (1).
21. The Knowledge Network for Biocomplexity (KNB) [Електронний ресурс] : [Сайт]. – Режим доступу : <https://knb.ecoinformatics.org>. – Назва з екрана.

## **ДОДАТКИ**



## ДОДАТОК А

### Критичні точки розподілу t-Ст'юдента

Число ступеню свободи, k	Рівень значущості (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	2	3	4	5	6	7
1	6,32	12,70	31,82	63,70	318,30	637,00
2	2,92	4,30	6,97	9,92	22,33	31,60
3	2,35	4,30	4,54	5,84	10,22	12,90
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,13	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,5
60	1,68	2,02	2,42	2,70	3,31	3,55
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	<b>0,05</b>	<b>0,025</b>	<b>0,01</b>	<b>0,005</b>	<b>0,001</b>	<b>0,0005</b>
	<b>Рівень значущості α (одностороння критична область)</b>					

## ДОДАТОК Б

Критичні значення F-критерію (при  $\alpha = 0,05$ )

Ступені свободи, $k_1$	Ступені свободи, $k_2$						
	1	2	3	4	5	6	50
9	5,12	4,26	3,86	3,63	3,48	3,37	2,8
10	4,96	4,10	3,71	3,48	3,33	3,22	-
11	4,84	3,98	3,59	3,36	3,20	3,09	2,5
12	4,75	3,88	3,49	3,26	3,11	3,00	-
13	4,67	3,80	3,41	3,18	3,02	2,92	2,32
14	4,60	3,74	3,34	3,11	2,96	2,85	-
15	4,54	3,68	3,29	3,06	2,90	2,79	2,18
16	4,49	3,63	3,24	3,01	2,85	2,47	-
17	4,45	3,59	3,20	2,96	2,81	2,70	2,08
18	4,41	3,56	3,16	2,93	2,77	2,66	-
19	4,33	3,52	3,13	2,90	2,74	2,63	2,08
20	4,35	3,49	3,10	2,87	2,71	2,90	-
21	4,32	3,47	3,07	2,84	2,68	2,57	1,93
22	4,30	3,44	3,05	2,82	2,66	2,55	-
23	4,28	3,42	3,03	2,80	2,64	2,53	1,88
24	4,26	3,40	3,01	2,78	2,62	2,51	-
25	4,24	3,38	2,99	2,76	2,60	2,49	1,84
26	4,22	3,37	2,98	2,74	2,59	2,47	-
27	4,21	3,35	2,96	2,73	2,57	2,46	1,80
28	4,20	3,34	2,95	2,71	2,56	2,44	-
29	4,18	3,38	2,93	2,70	2,54	2,43	1,77
30	4,17	3,32	2,92	2,69	2,53	2,42	-
35	4,11	3,26	2,86	2,63	2,48	2,36	-
40	4,08	3,23	2,84	2,61	2,45	2,34	-
50	4,03	3,18	2,79	2,56	2,40	2,29	-
60	4,00	3,15	2,76	2,52	2,37	2,25	-
$\infty$	3,84	2,99	2,60	2,37	2,21	2,09	-