

Міністерство освіти і науки України
Національний університет водного господарства
та природокористування

Кафедра геології та гідрології

01-05-303М

МЕТОДИЧНІ ВКАЗІВКИ

до виконання практичних робіт з дисципліни
**«СТАТИСТИЧНА ОБРОБКА ІНФОРМАЦІЇ
В НАУКАХ ПРО ЗЕМЛЮ»**

для здобувачів вищої освіти першого (бакалаврського) рівня
за освітньо-професійною програмою «Геологія»
спеціальності 103 «Науки про Землю»
денної та заочної форм навчання

Рекомендовано
науково-методичною радою
з якості ННІ ЕАВГ
Протокол № 4 від 17.12.2024 р.

Рівне – 2024

Методичні вказівки до виконання практичних робіт з дисципліни «Статистична обробка інформації в науках про Землю» для здобувачів вищої освіти першого (бакалаврського) рівня за освітньо-професійною програмою «Геологія» спеціальності 103 «Науки про Землю» денної форми навчання. [Електронне видання] / Косяк Д. С.– Рівне : НУВГП, 2024. – 27 с.

Укладач: Косяк Д. С., к.геогр.н., доцент кафедри геології та гідрології.

Рецензент: Холоденко В. С., к.геогр.н, доцент кафедри геології та гідрології.

Відповідальний за випуск: Мельничук В. Г., доктор геологічних наук, професор, завідувач кафедри геології та гідрології.

Керівник освітньої програми «Геологія» спеціальності 103 «Науки про Землю»: Мельничук В. Г., д. геол. н., професор, в.о. завідувача кафедри геології та гідрології.

© Д. С. Косяк, 2024

© НУВГП, 2024

Вступ.....	4
1. Класичне та статистичне визначення ймовірності геологічних даних	6
2. Вибірки та їх подання	9
3. Визначення статистичного зв'язку між геологічними ознаками за допомогою кореляції	14
4. Визначення статистичних оцінок вибірок геологічних даних.....	22
5. Оцінка однорідності рядів спостережень за статистичними критеріями	22
6. Оцінка однорідності рядів спостережень за параметричними критеріями.....	23
7. Оцінка однорідності рядів спостережень за непараметричними критеріями.....	23
Питання гарантованого рівня знань.....	24
Рекомендована та базова література.....	26
Допоміжна література.....	26

Вступ

Методичні вказівки призначені для виконання студентами практичних робіт з дисципліни «**Статистична обробка інформації в науках про Землю**». Вони складені відповідно до робочої програми дисципліни «Статистична обробка інформації в науках про Землю» та навчальних планів освітньо-професійної програми «Геологія» спеціальності 103 «Науки про Землю».

Дисципліна «Статистична обробка інформації в науках про Землю» є навчальною дисципліною з циклу обов'язкові компоненти освітньої програми «Геологія» за спеціальністю 103 «Науки про Землю» та невід'ємним складником формування професійної компетентності студентів.

Метою викладання дисципліни «Статистична обробка інформації в науках про Землю» є формування у майбутніх фахівців первинних знань із математичної статистики та статистичної обробки інформації в науках про Землю для вивчення кількісних і якісних характеристик природних явищ та процесів, а також для формулювання ними певних закономірностей, висновків та прогнозів.

Основними завданнями навчальної дисципліни є ознайомлення здобувачів з вибіркоким базовим понятійним та термінологічним апаратом математичної статистики, методами систематизації, обробки, аналізу інформації в науках про Землю, використання статистичних даних для наукових і практичних висновків та досліджень.

У результаті вивчення навчальної дисципліни студент повинен **знати**: основні поняття математичної статистики та статистичної обробки інформації в науках про Землю; схему застосування методів математичної статистики; послідовність статистичної обробки інформації у дослідженнях; особливості застосування методів математичної статистики в науках про Землю; основні поняття теорії ймовірностей та основних характеристик статистичного ряду спостережень; закон нормального розподілу неперервних величин; застосування статистичних методів розрахунків для дослідження геологічних та гідрологічних процесів та явищ; поняття забезпеченості, повторюваності та щільності розподілу випадкових величин; актуальність застосування різноманітних оцінок

регіональних змін клімату; а також сучасні методи оцінки змін водного стоку на прикладі річок з дослідженням часових рядів спостережень та оцінкою їх однорідності; застосування картографічних методів при обробці та аналізі ;гідрометеорологічної інформації; геостатистичний аналіз та просторову інтерполяцію у дослідженнях.

вміти: використовувати нормативні документи при виконанні статистичних розрахунків в науках про Землю, зокрема в геології та гідрології; застосовувати методи математичної статистики в науках про Землю; визначати основні характеристики статистичного ряду спостережень; визначати коефіцієнт варіації та асиметрії методом найбільшої правдоподібності, та методом моментів; коефіцієнт кореляції та регресії; розраховувати досліджувані статистичні величини при відсутності даних спостережень та при короткому ряді спостережень; оцінювати точність обчислення C_V і C_S , та визначити відносні похибки; розраховувати координати емпіричної та теоретичної кривих забезпеченості та будувати їх на сітчатці ймовірностей; визначати загальні показники детермінаційної складової часового ряду спостережень; виконувати оцінку трендів середнього значення в часі та оцінку однорідності рядів спостережень за статистичними параметричними та непараметричними критеріями.

1. Класичне та статистичне визначення ймовірності геологічних даних

Метою є вивчення основних понять дослідження геологічних даних, класичного та статистичного визначення ймовірності геологічних даних та одержання практичних навиків з методів їх визначення.

Завдання. 1. Розглянути та вивчити основні поняття дослідження геологічних даних. 2. Класичного та статистичного визначення ймовірності геологічних даних. 2. Виконати практичне завдання за варіантом, яке видане викладачем.

Методика виконання. В залежності від того, як отримані геологічні дані, всі вони поділяються на ті, які отримані в результаті польових та лабораторних спостережень [24].

Одна з найбільш важливих характеристик геологічних даних є їхня якість. До першого типу відносяться дані, які характеризують наявність або відсутність деякої властивості або якості досліджуваного об'єкту і носять назву *якісних даних*. Дані другого типу представлені числом і їх прийнято називати *кількісними даними*. Прикладом їх може бути набір зразків будь-якої породи, наприклад, граніту. Визначення об'єкту спостережень для відібрання геологічних даних залежить в кожному випадку від специфіки поставленого геологічного завдання, а саме від мети дослідження, і чіткого формулювання комплексу умов, які ставляться перед певним дослідженням.

Основна вимога до вихідної інформації – повнота даних. Під повнотою фактів розуміють: 1) повноту просторового охоплення явищ або процесів, елементів досліджуваного процесу; 2) повноту охоплення сторін явищ, тобто повноту вихідних даних щодо всіх істотних ознак та властивостей явищ; 3) повноту охоплення у часі. Тут слід передбачати наявність явищ за максимально тривалий час [2].

При дослідженні геологічних даних необхідно поєднувати застосування абсолютних і відносних показників, ізольоване застосування відносних показників від абсолютних може призвести до помилкових висновків, особливо при аналізі рядів динаміки. Тому комплексне використання абсолютних і відносних показників дає змогу поглибити аналіз явищ та процесів, вірно виявити

закономірності та особливості їх появи та розвитку, а також зробити правильні висновки.

Класичне визначення ймовірності геологічних даних.
 Класичне визначення ймовірності в застосуванні до складних задач природничо-наукового або інженерного характеру зазнає труднощів принципового характеру. Перш за все, в ряді випадків виникає питання знаходження правильного способу виділення «рівноможливих випадків».

Застосовується за таких умов: дослід (випробування) має скінченну кількість n рівноможливих взаємно виключних елементарних наслідків, а подія A відбувається з появою буд-якого з n (A) певних наслідків [4].

Наслідки, з яких складається подія A називаються сприятливими подіями A . Ймовірністю події A називається відношення кількості сприятливих події A наслідків до загальної кількості наслідків.

$$P(A) = \frac{n(A)}{n} \dots\dots\dots (1.1)$$

Кількість варіантів розміщення на n місцях k різних елементів визначаємо за формулою

$$A_n^k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!} \dots\dots\dots (1.2)$$

число розміщень з n по k .

де $n!$ – число розміщень n елементів на n місцях.

При цьому варіанти розміщення вважаються різними, якщо вони відрізняються зайнятими місцями, а також порядком, в якому розміщені на них елементи. На кожному місці може бути розміщено не більше одного елемента.

Кількість елементів вибору k елементів із загальної кількості n визначаємо за формулою

$$C_n^k = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \dots\dots\dots (1.3)$$

число комбінацій з n по $k!$. При цьому варіанти вибору вважаються різними лише в тому випадку, якщо вони відрізняються складом вибірок.

$$C_n^0 = C_n^n = 1, 0! = 1 \dots\dots\dots (1.4)$$

При $k) \frac{n}{2}$ для спрощення розрахунків корисно використовувати формулу

$$C_n^k = C_n^{n-k} \dots\dots\dots (1.5)$$

Якщо дослід складається з k окремих незалежних дослідів, то загальна кількість наслідків n_1, n_2, \dots, n_k , де n_i – кількість наслідків окремого дослідів.

Статистичне визначення ймовірності геологічних даних.

Випадковою називається змінна величина, яка в результаті дослідження може з'явитися, а може не з'явитися, незалежно від бажання дослідника. Поняття ймовірності є базовим поняттям у теорії ймовірностей і широко використовується в математичній статистиці. Ймовірність P появи події A називається відношення числа m випробувань, що сприяють появі події A , до загального числа n всіх можливих випробувань:

$$P(A) = \lim_{n \rightarrow \infty} \frac{m}{n} \dots \dots \dots (1.6)$$

Ймовірність змінюється від нуля (якщо $P(A)=0$, то подія неможлива) до одиниці (якщо $P(A)=1$, то подія достовірна).

Отже визначити ймовірність появи будь-якої події не викликає ніяких труднощів, якщо відомі значення всіх можливих подій та значення їх сприяючої частини. Існують умови, за яких значення m та n невідомі і безпосередньо не можуть бути знайдені.

Статистичне визначення є універсальним, але на практиці дає змогу визначити ймовірність лише наближено, та й то лише тоді, коли є можливість проводити достатню кількість випробувань. Тому існують визначення, які використовуються хоча і в окремих випадках, але завдяки яким можна розрахувати ймовірність точно та без випробувань: класичне та геометричне визначення ймовірності.

Статистичне означення ймовірності має описовий, а не формально-математичний характер. Проте це не зменшує його цінності, оскільки пізнання закономірностей ніколи не береться з нічого, йому завжди передують експеримент, спостереження.

Приклад 1. Знайти ймовірність того, що при розвідці корисних копалин на 6 ділянках випаде 2 парні ділянки з потрібною гірською породою.

Розв'язок.

Події $A_1, A_2, A_3, A_4, A_5, A_6$, де подія A_i – випало i ділянок, утворюють повну систему рівноможливих подій. Якщо подія A – випало 2 парні ділянки з потрібною породою, то їй сприяють події A_2, A_4, A_6 . Тому, за означенням

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

Приклад 2. Досліди виробки корисних копалин занумеровані в пробах з номерами від 1 до 100. Яка ймовірність того, що номер навмання вибраної проби не містить цифри 3, в якій міститься проба крейди?

Розв'язок.

Безпосереднім підрахунком знаходимо, що кількість чисел у межах першої сотні, які містять цифру 3, дорівнює 19. Це числа 3, 13, 23, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 43, 53, 63, 73, 83, 93. Тоді чисел, що не містять цифри 3, буде $100 - 19 = 81$. Тому за означенням

$$P(A) = \frac{81}{100} = 0,81.$$

2. Вибірки та їх подання

Метою є вивчення вибірок та їх подання.

Завдання. 1. Розглянути та вивчити основні поняття вибірки, репрезентативної вибірки, генеральної сукупності та їх подання. 2. Розглянути та вивчити основні поняття варіаційного ряду та частоти. 3. Виконати практичне завдання за варіантом, яке видане викладачем.

Методика виконання.

Метою статистичної обробки експериментальних даних є представлення їх у формі, зручній для аналізу та проведення такого аналізу з метою виявлення певних закономірностей у процесах, стан яких відображають експериментальні дані.

Генеральна сукупність – загальна кількість об'єктів дослідження, які мають деяку спільну ознаку, тобто **вибірка** – її частина, яка сформована певним науково обґрунтованим способом.

Число об'єктів вибіркової сукупності утворює обсяг вибірки.

Якщо властивості об'єктів вибірки правильно відображають властивості об'єктів генеральної сукупності, то кажуть, що вибірка **репрезентативна (представницька)**. У цьому випадку статистичні дані вважаються репрезентативними.

Статистичні дослідження проводять за певним планом, в якому можна виділити такі етапи:

1. Формулюють завдання дослідження та визначають обсяг потрібної вибірки (мета, об'єкти вивчення, їх кількість, які ознаки беруться до уваги, які характеристики об'єктів оцінюються).

2. Збирають потрібні дані, вибирають доцільну форму їх подання для подальшого дослідження, здійснюють їх наочне зображення після часткової обробки.

3. Проводять остаточну обробку статистичного матеріалу та його вивчення. За результатами формулюють певні висновки.

Елементи вибірки називають **варіантами** і записують їх у зростаючій послідовності (при дослідженні кількісної ознаки).

Сукупність варіант, записаних у порядку їх зростання, називають **варіаційним рядом**. Операцію запису варіант у порядку зростання називають **ранжуванням**, а ряд – **ранжувальним**. Варіаційними рядами користуються при невеликих об'ємах вибірки, а якщо вибірка велика, то її елементи групують в інтервали.

Коли ставиться питання про те, як часто зустрічаються ті чи інші дані, користуються терміном **частота**, розуміючи при цьому кількість даних, що трапляються. Частоту позначають буквою *m*. Відношення частоти (*m*) даних певного класу вибірки до кількості даних усієї вибірки (*n*) називають **відносною частотою** $\frac{m}{n}$ появи

варіанти.

Властивість відносних частот: сума відносних частот даних будь-якої вибірки дорівнює одиниці і для уникнення помилок ця умова завжди перевіряється.

Статистичним розподілом випадкової величини називається залежність значень варіант від їх відносних частот.

Приклад 1. Результати 20 спостережень ($n=20$) над кількісною варіантою зведено в таблицю: Визначити відносні частоти спостережуваної варіанти та записати статистичний розподіл у вигляді таблиці 2.1.

Таблиця 2.1

Результати спостережень над кількісною варіантою

Варіанта, <i>X</i>	2	6	10	12	14
Частота, <i>m</i>	1	5	7	3	4

Розв'язок.

Визначимо відносні частоти спостережуваної варіанти:

$$\frac{m_1}{n} = \frac{1}{20}; \frac{m_2}{n} = \frac{5}{20}; \frac{m_3}{n} = \frac{7}{20}; \frac{m_4}{n} = \frac{3}{20}; \frac{m_5}{n} = \frac{4}{20}. \quad (2.1)$$

Запишемо статистичний розподіл у вигляді таблиці 2.2.

Таблиця 2.2

Статистичний розподіл спостережень

Варіанта, X	2	6	10	12	14
Відносна частота, $\frac{m}{n}$	$\frac{1}{20}$	$\frac{5}{20}$	$\frac{7}{20}$	$\frac{3}{20}$	$\frac{4}{20}$

Виконуємо перевірку: сума відносних частот даних будь-якої вибірки дорівнює одиниці.

$$\frac{1}{20} + \frac{5}{20} + \frac{7}{20} + \frac{3}{20} + \frac{4}{20} = \frac{20}{20} = 1 \quad (2.2)$$

При великих n або неперервному розподілі ознаки X доцільно значення ознаки X групувати в інтервали і будувати так звану гістограму графічного зображення статистичного розподілу. Де на осі OX відкладають значення варіант X , а на осі OY – відносні частоти ω_i . Тоді точки (X_i, ω_i) сполучають відрізками, і отримують полігон відносних частот. Аналогічно одержують полігон частот, якщо по осі OY відкладають частоти.

Приклад. 2. В результаті вимірювання витрат води в річці (m^3/c) отримані значення у місцях спостережень групами студентів зведені у таблицю 2.3.

Таблиця 2.3

Вимірювання витрат води в річці (ранжувальний ряд)

1 група	2 група	3 група	4 група	5 група	6 група
79,95	80,14	80,14	80,18	80,20	80,24
80,24	80,28	80,28	80,29	80,30	80,31
80,32	80,33	80,33	80,35	80,36	80,36
80,37	80,37	80,38	80,38	80,38	80,38
80,41	80,42	80,42	80,43	80,43	80,46
80,46	80,47	80,47	80,48	80,48	80,48
80,49	80,50	80,50	80,51	80,54	80,54
80,55	80,56	80,56	80,56	80,58	80,58

80,60	80,64	80,64	80,66	80,66	80,69
80,70	80,72	80,72	80,72	80,73	80,73

Для побудови варіаційного інтервального ряду, візьмо 7 інтервалів, тоді довжина кожного інтервалу h буде дорівнювати

$$h = \frac{x_{\max} - x_{\min}}{7} = \frac{80,73 - 79,95}{7} \approx 0,12. \quad (2.3)$$

Визначимо межі інтервалів α_i та x_i як середнє арифметичне кінців інтервалів $x_i = \frac{\alpha_i + \alpha_{i+1}}{2}, i = 1, 2, 3, 4, 5, 6, 7,$ (2.4) де α_i - кінці інтервалів. У нашому випадку $i=1, 8$.

За частоту n_i беремо кількість значень ранжувального ряду (табл. 2.3), що попадають в i -ий інтервал. Якщо обидва кінці інтервалу збігаються із значеннями ряду (табл. 2.3), то до інтервалу зараховують лівий кінець.

Для подальших розрахунків потрібні відносні частоти (ω_i) та щільності відносних частот (P_i): $\omega_i = \frac{n_i}{60}; P_i = \frac{\omega_i}{h}, i=1, 7$ (2.5)

Одержані результати зводимо у таблицю 2.4, яку називають **варіаційним інтервальним рядом** досліджуваної випадкової величини.

Таблиця 2.4

Варіаційний інтервальний ряд вимірних витрат води

Інтервали	79,95 ÷ 80,07	80,07 ÷ 80,19	80,19 ÷ 80,31	80,31 ÷ 80,43	80,43 ÷ 80,55	80,55 ÷ 80,67	80,67 ÷ 80,79
x_i	80,01	80,13	80,25	80,37	80,49	80,61	80,73
n_i	1	3	10	17	15	10	4
ω_i	$\frac{1}{60}$	$\frac{3}{60}$	$\frac{10}{60}$	$\frac{17}{60}$	$\frac{15}{60}$	$\frac{10}{60}$	$\frac{4}{60}$
P_i	0,14	0,42	1,40	2,36	2,08	1,40	0,56

За даними таблиці 2.4 будемо полігон відносних частот (рис. 2.1).

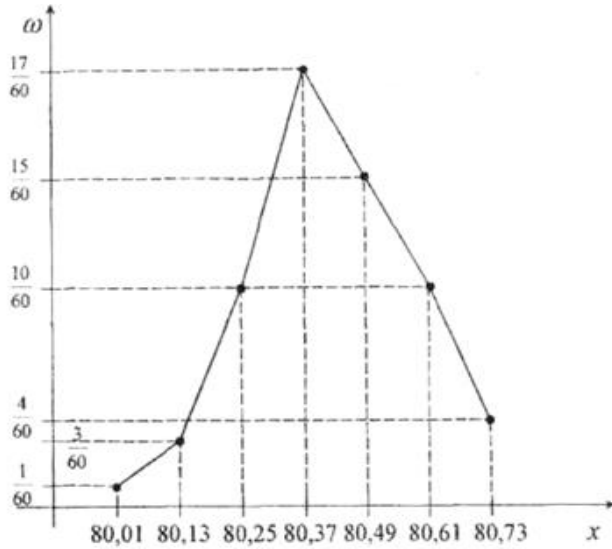


Рис. 2.1. Графік полігону відносних частот

За даними таблиці 2.4 будуємо гістограму щільності відносних частот (рис. 2.2).

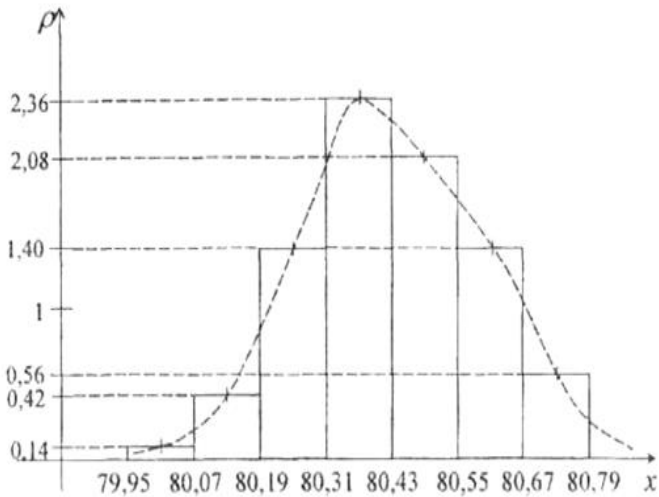


Рис. 2.2. Гістограма щільності відносних частот

3. Визначення статистичного зв'язку між геологічними ознаками за допомогою кореляції

Метою є визначення статистичного зв'язку між геологічними ознаками за допомогою кореляції.

Завдання. 1. Розглянути та вивчити основні поняття кореляції та її форми, напрям та степінь кореляційних зв'язків. 2. Коефіцієнт кореляції, рівняння регресії, їх застосування. 3. Виконати практичне завдання за варіантом, яке видане викладачем.

Методика виконання.

Часто доводиться мати справу з більш складними залежностями, ніж функціональна, наприклад, зв'язок між геологічними породами та їх межею, або опадами та урожаєм, або зв'язок між товщиною снігу зимою і об'ємом повені навесні. Тут кожному значенню однієї величини відповідає множина можливих значень іншої величини. Подібні залежності називають **кореляційними**.

Щоб визначити залежність між змінними X (геологічна порода) та Y (геологічна межа), оцінити їх напрям та інтенсивність. Порівнюючи різні види зв'язків, можна виділити три основних типи залежностей між змінними X та Y .

1. Функціональна залежність визначає значення змінної Y від X однозначно.

2. Кореляційна залежність визначає середнє значення змінної Y від X .

3. Стохастична залежність визначає розподіл змінної Y від X .

Найбільш загальною вважається стохастична залежність. Кореляційна залежність є стохастичною. Функціональна залежність розглядається як окремий випадок кореляційної залежності.

Кореляція (від лат. означає співвідношення) – це статистична залежність між випадковими величинами, що носить імовірнісний характер, тобто не має строго функціонального характеру.

Кореляційна залежність виникає тоді, коли одна з величин залежить не тільки від заданої другої, а й від деяких випадкових факторів; або, коли серед умов, від яких залежать обидві величини, є загальні для них обох.

За формою кореляція може бути **прямолінійною і криволінійною**, за напрямком – **прямою і оберненою**. При додатній кореляції залежність між величинами буде прямою: при збільшенні

однієї величини, збільшується й інша. При від'ємній кореляції залежність обернена: збільшення однієї величини пов'язано зі зменшенням другої. Нульовою називається кореляція за відсутністю зв'язку змінних. Даний вид кореляції може свідчити про відсутність лінійної залежності, а не взагалі про відсутність будь-якого статистичного зв'язку. Степінь кореляції вимірюється різними показниками зв'язку. Такими показниками є коефіцієнт кореляції, кореляційне відношення та ін.

Кореляційний зв'язок – це не точна залежність однієї величини від іншої.

Числовим значенням однієї змінної ставиться у відповідність середнє декількох значень інших. Наприклад, між кількістю внесених на поле добрив і врожайністю пшениці існує незаперечна залежність. Але це не означає, що конкретній кількості добрив відповідає визначена величина урожаю. На урожай впливає багато інших факторів: склад і структура ґрунту, різні методи посіву і таке інше.

Кореляційний зв'язок виявляється у середньому для усієї сукупності спостережень. По відношенню ж до окремих спостережень цей зв'язок є дуже неповним і неточним. Відомо, наприклад, що існує кореляція між вагою тварини і її висотою. Це означає, що більш високі тварини звичайно важчі за низьких. Та в деяких випадках низька тварина може виявитися важчою за високу.

Кореляційний зв'язок може мати різну степінь – від повної незалежності до функціональної залежності. Крім того, характер зв'язку між різними величинами може бути різний. Тому виникає необхідність визначити форму, напрям і степінь кореляційних зв'язків.

Кореляційне відношення визначається нижче.

Наприклад, при розрахунку та визначенні статистичного зв'язку між геологічними ознаками, річного стоку, агрометеорологічних явищ тощо, з коротким рядом спостережень можна використовувати два методи:

1. Безпосереднього приведення середньоарифметичної величини розрахункової певної геологічної ділянки, річки, площ урожаю і т. і. до багаторічного періоду на основі рівняння прямолінійної регресії.

2. Продовження ряду спостережень певної розрахункової величини.

В практичній роботі розрахунки при короткому ряді спостережень проводимо другим методом.

Короткий ряд продовжують шляхом встановлення аналітичного, або графічного зв'язку, наприклад візьмемо, між річним стоком у розрахунковому створі зі стоком річки-аналога. Даний метод також може використовуватися для дослідження будь-яких інших величин (геологічних, агрометеорологічних тощо).

Графіки зв'язку між величинами стоку розрахункової річки і річки-аналога будують за всі роки паралельних спостережень. Графіки зв'язку використовуються для продовження ряду і визначення наближеного значення коефіцієнта кореляції.

Для аналітичного способу розрахунку використовують рівняння прямолінійної регресії, що розраховується за роки спільних спостережень

$$Y_i - Y_0 = r \frac{\sigma_y}{\sigma_x} (X_i - X_0), \quad (3.1)$$

де X_i і Y_i - значення річного стоку річки-аналога та в розрахунковому створі за i -ті роки; X_0 і Y_0 - відповідно середні багаторічні величини річного стоку за період паралельних спостережень в створі річки-аналога і розрахунковому створі; σ_x і σ_y - середні квадратичні відхилення рядів X і Y ; r - коефіцієнт кореляції, що характеризує тісноту зв'язку між стоком розрахункової річки і річки-аналога.

Коефіцієнт кореляції r розраховується за формулою

$$r = \frac{\sum(X_1 - X_0) \cdot (Y_1 - Y_0)}{\sqrt{\sum(X_1 - X_0)^2 \cdot \sum(Y_1 - Y_0)^2}}, \quad (3.2)$$

а σ_x і σ_y розраховуються за формулами

$$\sigma_y = \sqrt{\frac{\sum(Y_1 - Y_0)^2}{n-1}}, \quad (3.3)$$

$$\sigma_x = \sqrt{\frac{\sum(X_1 - X_0)^2}{n-1}}, \quad (3.4)$$

Зв'язок буде вважається надійним, коли значення коефіцієнта кореляції більше 0,70. Коефіцієнт кореляції розраховують, коли кількість років паралельних спостережень більша 15. Коли кількість років менша, імовірна похибка коефіцієнту кореляції буде великою, а зв'язок маловірогідним.

Приклад. 1. Дано: Середньорічні витрати р. Турія біля с. Бузакі за період з 1962 по 1981 рік; середньобагаторічна величина $Q_{сер} = 10,05$ м³/с; коефіцієнт варіації річного стоку $C_V = 0,76$; відносна похибка обчислення середнього багаторічного значення $\sigma_Q = 17,05\%$.

Необхідно: Визначити норму річного стоку р. Турія біля с. Бузаки, як при короткому ряді спостережень.

Розв'язок. За річку-аналог приймаємо р. Турія біля м. Ковеля із періодом спостережень в 46 років. Норма річного стоку її складає 3,78 м³/с. Період спільних спостережень в обох створах складає 20 років.

Згідно таблиці 3.1:

$$X_0=87,97/20=4,40 \text{ м}^3/\text{с},$$
$$Y_0=201,06/20=10,05 \text{ м}^3/\text{с}.$$

Обчислення коефіцієнта кореляції зв'язку річного стоку р. Турія біля с. Бузаки та м. Ковеля наведені в таблиці 3.1.

Коефіцієнт кореляції зв'язку обчислюємо за формулою (3.2)

$$r = \frac{137,6}{\sqrt{75,6 \cdot 338,2}} = 0,86.$$

Середні квадратичні відхилення σ_n і σ_{na} обчислюємо за формулами 3.3, 3.4.

$$\sigma_y = \sqrt{\frac{338,2}{20 - 1}} = 4,21;$$
$$\sigma_x = \sqrt{\frac{75,61}{20 - 1}} = 1,99.$$

Підставивши всі параметри в формулу 3.1 отримаємо рівняння регресії:

$$Y_1 - 10,05 = 0,86 * (4,21/1,99) * (X_1 - 4,40)$$
$$Y_1 = 1,82 * X_1 + 2,04$$

За отриманим рівнянням продовжують період спостережень на розрахунковій річці за роки в які є спостереження на річці аналогу (X_1), а відсутні в розрахунковому створі, і розрахунок повторюють для продовженого ряду, аналогічно таблиці 3.2.

Розрахунок основних параметрів рівняння регресії для прикладу з річним стоком річки можна звести у форму таблиці 3.1. Аналогічну форму можна використовувати і для інших прикладів.

Таблиця 3.1

Обчислення коефіцієнта кореляції зв'язку річного стоку р. Турія біля м. Ковеля та с. Бузаки при короткому ряді спостережень

№ з/п	Рік	Річний стік р.Турії, м ³ /с		$X_i - X_{\text{ср}}$	$(X_i - X_{\text{ср}})^2$	$Y_i - Y_{\text{ср}}$	$(Y_i - Y_{\text{ср}})^2$	$(X_i - X_{\text{ср}})(Y_i - Y_{\text{ср}})$
		біля м.Ковеля, X_i	біля с.Бузакки, Y_i					
1	2	3	4	5	6	7	8	9
1	1962	3,82	7,20	-0,58	0,34	-2,85	8,14	1,65
2	1963	1,82	4,01	-2,58	6,65	-6,04	36,52	15,58
3	1964	2,15	4,22	-2,25	5,06	-5,84	34,02	13,12
4	1965	4,09	10,20	-0,31	0,10	0,15	0,02	-0,05
...
20	1981	5,18	14,20	0,78	0,61	4,15	17,20	3,24
Сума		87,98	201,06	0,00	75,61	0,00	338,20	137,60
Середнє		4,40	10,05					

Розрахунок для продовженого ряду спостережень виконують у формі таблиці 3.2.

Приклад 2. Нехай є дві випадкові величини X (геологічна порода – крейда або інша порода) і Y (геологічна межа крейди у різних свердловинах, або межа іншої породи), які знаходяться в кореляційній залежності, бо кожному значенню однієї з них відповідає певний розподіл імовірностей другої величини.

Нехай у результаті n випробувань одержані пари значень випадкових величин X і Y : (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) . Точки x і y групуються навколо деякої прямої $y = kx + b$ (рис. 3.1). Виникає задача: знайти таку пряму $y = kx + b$, щоб величина

$$d = \sum_{i=1}^n (y_i - kx_i - b)^2 \text{ була найменшою.}$$

Таблиця 3.2

Приклад розрахунку норми річного стоку р. Турії для продовженого ряду спостережень

№ з/п	Роки	$Q_i, \text{ м}^3/\text{с}$	$Q_i, \text{ м}^3/\text{с}$ в ранжовано- му ряді	$K_i = Q_i/Q_0$	$K_i - 1$	$(K_i - 1)^2$	$(K_i - 1)^3$	$\lg K_i$	$K_i \lg K_i$	$P = \frac{n * 100\%}{n + 1}$
1	2	3	4	5	6	7	8	9	10	11
1	1946	2,42	9,46	2,51	1,51	2,27	3,41	0,40	1,00	2,13
2	1947	2,43	8,33	2,21	1,21	1,46	1,76	0,34	0,76	4,26
3	1948	8,33	7,69	2,04	1,04	1,08	1,11	0,31	0,63	6,38
4	1949	3,56	7,64	2,02	1,02	1,05	1,07	0,31	0,62	8,51
...
43	1988	5,67	1,81	0,48	-0,52	0,27	-0,14	-0,32	-0,15	91,49
44	1989	3,85	1,79	0,47	-0,53	0,28	-0,15	-0,32	-0,15	93,62
45	1990	2,90	1,16	0,31	-0,69	0,48	-0,33	-0,51	-0,16	95,74
46	1991	3,64	1,09	0,29	-0,71	0,51	-0,36	-0,54	-0,16	97,87
Сума		173,67	173,67	46,00	0,00	12,04	6,48	-2,46	2,42	

Розв'язок. За допомогою методу найменших квадратів можна показати, що величина d буде найменшою при таких значеннях k і b :

$$k = \frac{\bar{x}\bar{y} - \bar{x}\cdot\bar{y}}{\bar{x}^2 - (\bar{x})^2}, \quad b = \bar{y} - k\bar{x}, \quad (3.5)$$

де $\bar{x}\bar{y} = \frac{1}{n} \sum_{i=1}^n x_i y_i$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, $\bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$,

$$\text{Тоді рівняння буде мати вигляд } y - \bar{y} = k(x - \bar{x}), \quad (3.6)$$

Пряма, задана рівнянням (3.6), називається **прямою вибіркової регресії Y на X** .

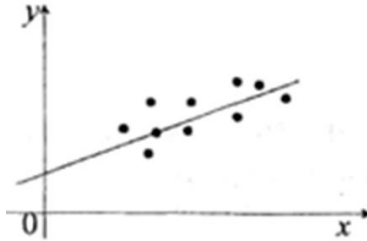


Рис. 3.1. Графік, на якому точки x і y групуються навколо деякої прямої $y = kx + b$

Приклад 3. Нехай в результаті 10 геологічних випробувань отримані такі пари значень випадкових величин X і Y : (1, 1), (2, 3), (2, 4), (3, 6), (4, 5), (6, 3), (7, 5), (8, 4), (9, 12), (10, 10).

Записати рівняння прямої вибіркової регресії Y на X .

Розв'язок. Для обчислення чисел k і b складемо розрахункову таблицю 3.3.

Таблиця 3.3

Вихідні дані для обчислення чисел k і b

x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	1	1	1	1
2	3	4	6	9
2	4	4	8	16
3	6	9	18	36
4	5	16	20	25
6	3	36	18	9
7	5	49	35	25
8	4	64	32	16
9	12	81	108	144
10	10	100	100	100
5,2	5,3	36,4	34,6	38,1

У нижньому рядку таблиці 3.3 записані середні значення $\bar{x}, \bar{y}, \bar{x}^2, \bar{x}\bar{y}, \bar{y}^2$. За формулами 3.5 отримаємо

$$k = \frac{34,6 - 5,2 \cdot 5,3}{36,4 - (5,2)^2} = \frac{7,04}{9,36} \approx 0,752,$$

звідки шукане рівняння прямої регресії Y на X має вигляд $y - 5,3 = 0,752(x - 5,2)$.

В теорії кореляції степінь лінійної залежності між випадковими величинами X та Y характеризується кореляційним моментом μ_{xy} і коефіцієнтом кореляції r_{xy} , які пов'язані між собою рівністю

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}. \quad (3.7)$$

Статистичною оцінкою цих параметрів є відповідно вибірковий кореляційний момент $\tilde{\mu}_{xy}$ і вибірковий коефіцієнт кореляції \tilde{r}_{xy} , які обчислюються за формулами

$$\tilde{\mu}_{xy} = \bar{x}\bar{y} - \bar{x} \cdot \bar{y}; \tilde{r}_{xy} = \frac{\bar{x}\bar{y} - \bar{x} \cdot \bar{y}}{\sqrt{\bar{x}^2 - (\bar{x})^2} \cdot \sqrt{\bar{y}^2 - (\bar{y})^2}}. \quad (3.8)$$

Можна показати, що $-1 \leq \tilde{r}_{xy} \leq 1$, причому значення -1 і 1 коефіцієнт кореляції \tilde{r}_{xy} приймає лише тоді, коли всі точки множини $\{(x_i, y_i), i = \bar{1}, \bar{n}\}$ розміщені на прямій $y = kx + b, k \neq 0$.

Для незалежних випадкових величин X і Y числа $\tilde{\mu}_{xy}$ і \tilde{r}_{xy} близькі до нуля. Отже, за величиною \tilde{r}_{xy} можна судити про степінь лінійної залежності між випадковими величинами X і Y : якщо \tilde{r}_{xy} близьке до ± 1 , то залежність між X і Y близьке до лінійної; якщо \tilde{r}_{xy} близьке до нуля, то залежність між X і Y – статистична.

Для прикладу 3 число $\tilde{r}_{xy} = \frac{34,6 - 5,2 \cdot 5,3}{\sqrt{9,36} \cdot \sqrt{10,01}} = 0,797$.

Це значить, що степінь лінійної залежності між випадковими величинами X і Y досить висока.

В складніших випадках залежність між випадковими величинами X і Y може бути квадратичною $y = ax^2 + bx + c$, показниковою $y = a \cdot b^x$ тощо. В усіх цих випадках також застосовують метод найменших квадратів для обчислення параметрів a, b, c .

4. Визначення статистичних оцінок вибірок геологічних даних

Метою є визначення статистичних оцінок вибірок геологічних даних.

Завдання. 1. Виконати оцінку однорідності двох вибірок експериментальних даних за модифікацією Сіджела-Тьюкі критерію Вілкоксона. 2. Результати представити у табличній формі.

Методика виконання.

Оцінку однорідності двох вибірок експериментальних даних за модифікацією Сіджела-Тьюкі критерію Вілкоксона виконуємо згідно Методичних вказівок (01-05-67) до виконання самостійної роботи з дисципліни «Статистична обробка інформації в науках про Землю» для здобувачів вищої освіти першого (бакалаврського) рівня за освітніми програмами «Геологія», «Гідрологія» спеціальності 103 «Науки про Землю» та освітніми програмами «Конструктивна географія, управління водними та мінеральними ресурсами», «Географія рекреації та туризму» спеціальності 106 «Географія», розділ 4, стор.20.

5. Оцінка однорідності рядів спостережень за статистичними критеріями

Метою є оцінка однорідності рядів спостережень за статистичними критеріями.

Завдання. 1. Виконати оцінку однорідності двох вибірок експериментальних даних за серійним критерієм. 2. Результати представити у табличній формі.

Методика виконання.

Оцінку однорідності двох вибірок експериментальних даних за серійним критерієм виконуємо згідно Методичних вказівок (01-05-67) до виконання самостійної роботи з дисципліни «Статистична обробка інформації в науках про Землю» для здобувачів вищої освіти першого (бакалаврського) рівня за освітніми програмами «Геологія», «Гідрологія» спеціальності 103 «Науки про Землю» та освітніми програмами «Конструктивна географія, управління водними та мінеральними ресурсами», «Географія рекреації та туризму» спеціальності 106 «Географія», розділ 5, стор.26.

6. Оцінка однорідності рядів спостережень за параметричними критеріями

Метою є оцінка однорідності рядів спостережень за параметричними критеріями.

Завдання. 1. Виконати оцінку однорідності двох вибірок експериментальних даних за критерієм Колмогорова-Смірнова. 2. Результати представити у табличній формі.

Методика виконання.

Оцінку однорідності двох вибірок експериментальних даних за критерієм Колмогорова-Смірнова виконуємо згідно Методичних вказівок (01-05-67) до виконання самостійної роботи з дисципліни «Статистична обробка інформації в науках про Землю» для здобувачів вищої освіти першого (бакалаврського) рівня за освітніми програмами «Геологія», «Гідрологія» спеціальності 103 «Науки про Землю» та освітніми програмами «Конструктивна географія, управління водними та мінеральними ресурсами», «Географія рекреації та туризму» спеціальності 106 «Географія», розділ 6, стор.30.

7. Оцінка однорідності рядів спостережень за непараметричними критеріями

Метою є оцінка однорідності рядів спостережень за непараметричними критеріями.

Завдання. 1. Виконати оцінку однорідності двох вибірок експериментальних даних за критерієм знаків. 2. Виконати оцінку однорідності двох вибірок експериментальних даних за критерієм Вілкоксона для зв'язаних пар. 3. Результати представити у табличній формі.

Методика виконання.

Оцінку однорідності двох вибірок експериментальних даних за критерієм знаків та оцінку однорідності двох вибірок експериментальних даних за критерієм Вілкоксона для зв'язаних пар виконуємо згідно Методичних вказівок (01-05-67) до виконання самостійної роботи з дисципліни «Статистична обробка інформації в науках про Землю» для здобувачів вищої освіти першого

(бакалаврського) рівня за освітніми програмами «Геологія», «Гідрологія» спеціальності 103 «Науки про Землю» та освітніми програмами «Конструктивна географія, управління водними та мінеральними ресурсами», «Географія рекреації та туризму» спеціальності 106 «Географія», розділ 7.1, стор.36.

Питання гарантованого рівня знань

1. Поняття статистики, математичної статистики та статистичної обробки інформації.

2. Описова статистика.

3. Статистичні висновки.

4. Випадкова змінна величина.

5. Поняття забезпеченості, повторюваності та щільності розподілу випадкових величин.

6. Статистична ймовірність.

7. Дискретна величина.

8. Неперервна величина.

9. Основні характеристики статистичного ряду.

10. Математичне сподівання. Середнє квадратичне відхилення.

11. Дисперсія. Мода

12. Розмах вибірки. Ексцес вибірки.

13. Коефіцієнт асиметрії та варіації.

14. Генеральна сукупність та частота.

15. Вибірка та репрезентативна вибірка, їх подання.

16. Варіаційний ряд та статистичний розподіл. Класичне та статистичне визначення ймовірності в науках про Землю.

17. Закон нормального розподілу.

18. Розподіли "хі-квадрат", t-розподіл Стьюдента, F-розподіл Фішера-Снедекора.

19. Гідрологічні розрахунки та визначення норми річного стоку.

20. Визначення коефіцієнта варіації та асиметрії методом найбільшої правдоподібності.

21. Визначення коефіцієнта варіації та асиметрії методом моментів.

22. Розрахунок норми річного стоку при відсутності даних спостережень.

23. Обчислення статистичних параметрів річного стоку річки в табличній формі.

24. Основні вимоги до річок-аналогів.

25. Кореляція. Кореляційний зв'язок. Форми кореляції. Коефіцієнт кореляції. Визначення статистичного зв'язку між показниками в науках про Землю за допомогою кореляції.

26. Методи розрахунку норми річного стоку річки при короткому ряді спостережень.

27. Оцінка точності обчислення C_v і C_s . Визначення відносних похибок.

28. Розрахунок координат емпіричної та теоретичної кривих забезпеченості з нанесенням їх на сітчатку ймовірностей.

29. Актуальність застосування різноманітних оцінок регіональних змін клімату.

30. Сучасні методи оцінки змін водного стоку річок та методика їх досліджень.

31. Загальні показники детермінаційної складової часового ряду спостережень.

32. Оцінка трендів середнього значення в часі.

33. Оцінка однорідності рядів спостережень за статистичними критеріями в науках про Землю. Параметричні та непараметричні критерії.

34. Застосування картографічних методів при обробці та аналізі інформації.

35. Картографічний та картометричний методи досліджень.

36. Графічний та математико-статистичний аналізи.

37. Просторова інтерполяція у дослідженнях в науках про Землю.

38. Схема застосування методів математичної статистики.

39. Послідовність статистичної обробки інформації у дослідженнях.

40. Особливості застосування методів математичної статистики в науках про Землю (генетичний метод та інші).

Рекомендована та базова література

Базова

1. Мельниченко О. П., Якименко І. Л., Шевченко Р. Л. Статистична обробка експериментальних даних : навч. посіб. Біла Церква, 2006. 34 с.
2. Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань) : навч. посіб. К. : Центр учбової літератури, 2012. 448 с.
3. Руденко В. М. Математична статистика : навч. посіб. К. : Центр учбової літератури, 2012. 304 с.
4. Омельченко В. Г., Здерка Т. В., Куровець С. С. Математична статистика та обробка геологічної інформації : лаб. Практикум/ Івано-Франківськ : ІФНТУНГ, 2011. 57 с.

Допоміжна

1. Самойленко В. М. Географічні інформаційні системи та технології : підручник. К. : Ніка-Центр, 2010. 448 с.
2. Геоінформаційні системи в науках про Землю : монографія / Зацерковний В. І., Тишаєв І. В., Віршило І. В., Демидов В. К. Ніжин : НДУ ім. М. Гоголя, 2016. 510 с.
3. Галік О. І., Сливка П. Д. Методичні вказівки (075-84). Рівне : НУВГП, 2004. 28 с.
4. Бузіна І. М., Литвиненко Ю. О. Земельно-кадастрове картографування: Використання карт земельних ресурсів : конспект лекцій. Харк. нац. аграр. ун-т. Х., 2016. 132 с.
5. Гребінь В. В. Сучасний водний режим річок України (ландшафтно-гідрологічний аналіз). К. : Ніка-Центр, 2010. 316 с.
6. Горбачова Л. О., Бібик Т. О. Часова однорідність характеристик водного стоку в басейні річки Боржава/ *Наук. праці УкрНДГМІ*. 2012. Вип. 262.
7. Горбачова Л. О., Баужа Т. О. Багаторічні коливання середньорічних витрат води на річках і струмках Закарпатської воднобалансової станції/ *Гідрологія, гідрохімія, гідроекологія: Матеріали п'ятої Всеукр. наук. конф. Чернівці* : Чернівецький нац. ун-т, 2011. С. 52-54.
8. Горбачова Л. О., Кошкіна О. В. Часові закономірності дат настання основних характеристик весняного водопілля в басейні р. Десна/ *Гідрологія, гідрохімія і гідроекологія*. 2013. Т. 2 (29). С. 30–37.

9. Клімат України / Дячука В. А., Бабіченко : за ред. В. М. Ліпінського. К. : Видавництво Раєвського, 2003. 343 с.
10. Гидрологические и водно-балансовые расчеты / Под ред. Н. Г. Галушенко. К. : Вища шк. Головное изд-во, 1987. 248 с.
11. Агрометеорологічні прогнози. Практикум : навчальний посібник / Божко Л. Ю., Барсукова О. А. Одеса, 2011. 229 с.
12. Світличний О. О., Плотницький С. В. Основи геоінформатики : навчальний посібник / За ред. О. О. Світличного. Суми : ВТД «Університетська книга», 2006. 295 с.
13. Кіндюк Б. В. Гідрографічна мережа та зливовий стік річок Українських Карпат : автореф. дис. ... д-ра геогр. наук: спец. 11.00.07 / КНУ ім. Тараса Шевченка. К., 2004. 30 с.
14. Струтинська В. М. Вплив змін клімату на термічний та льодовий режими річок басейну Дніпра (в межах України) з другої половини ХХ ст.: автореф. дис. ... канд. геогр. наук: спец. 11.00.07 / КНУ ім. Тараса Шевченка. К., 2008. 20 с.
15. Василенко Є. В. Характеристики весняного водопілля правобережних приток р. Прип'ять в сучасних кліматичних умовах: автореф. дис. ... канд. геогр. наук: спец. 11.00.07 / Київ. 2012. 20 с.
16. Косяк Д. С., Холоденко В. С., Галік О. І., Будз О. П. Гідрометрія: практикум : навч. посіб. Рівне : НУВГП, 2018. 254 с. URL: <http://ep3.nuwm.edu.ua/11563/>
17. Сливка П. Д., Новосад Я. О., Будз О. П. Гідрологія та регулювання стоку. Рівне : УДУВГП, 2003. 310 с. URL: <http://ep3.nuwm.edu.ua/157/1/01-03-22.pdf>
18. Сливка П. Д. Водогосподарські розрахунки : навч. посіб. Рівне : НУВГП, 2010. 78 с. URL: <http://ep3.nuwm.edu.ua/id/eprint/1648>
19. Будз О. П. Гідрологія : інтерактивний комплекс навчально-методичного забезпечення. Рівне : НУВГП, 2008. 168 с. URL: <http://ep3.nuwm.edu.ua/id/eprint/1842>
20. Брушковський О. Л., Дубчак І. В. Теорія ймовірностей і математична статистика : навчальний посібник для студентів I курсу заочної форми навчання напрямів підготовки 6.030504 "Економіка підприємства", 6.030509 "Облік і аудит", 6.030508 "Фінанси і кредит". Рівне : НУВГП, 2010. 117 с. URL: <http://ep3.nuwm.edu.ua/4535/>