

# Лабораторна робота №1

## Статистична обробка геологічної інформації

### 1. Мета роботи

Метою лабораторної роботи є освоєння методики та прийомів первинної статистичної обробки геологічних даних, що включають побудову гістограм, обчислення статистичних характеристик та вивчення характеру статистичного зв'язку між різними фізичними параметрами – вмістом магнетиту і щільністю, визначеними за зразками гірських порід.

В результаті виконання даної роботи студент повинен: знати методику найпростішої статистичної обробки петрофізичних параметрів; вміти вручну обчислювати статистичні характеристики вибірки, будувати гістограми, поле кореляції і визначати ступінь тісноти кореляційного зв'язку двох параметрів; розуміти зміст статистичних характеристик вибірки, необхідність проведення статистичного аналізу результатів визначення фізичних властивостей за зразками гірських порід.

### 2. Теоретичні основи

Статистична обробка вимірів служить для кількісної оцінки закономірностей, виявлених при аналізі результатів вивчення фізичних властивостей гірських порід і корисних копалин.

Статистична обробка матеріалів складається з двох етапів: 1) попередньої обробки, головною метою якої є складання вихідної документації; 2) математичної обробки, що включає виділення петрофізичних груп, побудову кривих розподілу, розрахунок статистичних параметрів, визначення характеру зв'язку між властивостями. Попередня обробка полягає, головним чином, у складанні каталогів даних про склад і фізичні властивості гірських порід, планів і розрізів з нанесенням точок відбору і величини фізичного параметра. Математична обробка петрофізичних даних дозволяє встановити основні закономірності зміни фізичних параметрів досліджуваних гірських порід. У даній роботі передбачається проведення математичної обробки петрофізичних даних, а результати первинної обробки надаються у якості вихідних даних.

**Обчислення статистичних характеристик і побудова гістограм.** На практиці результати експерименту представляють в вигляді деякої випадкової величини  $X$ . Наприклад, вимірювання фізичних властивостей гірських порід (щільність, вміст магнетиту, питомий електричний опір та ін.) є спостереженнями випадкових величин. Іншими словами, значення конкретного фізичного параметра є випадковою величиною. За сукупністю величин визначаються статистичні характеристики, що дозволяють судити про закономірності розподілу цієї випадкової величини.

Якщо проведено  $N$  вимірювань деякого фізичного параметра гірської породи, то говорять про сукупність вимірювань даного параметру  $X(x_1, x_2, \dots, x_i, \dots, x_N)$ .

При вимірах однакової точності найбільш вірогідним значенням вимірюваної величини є середнє арифметичне сукупності

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i,$$

яке характеризує стан випадкової величини на числовій осі.

Як показник величини відхилення окремих значень від середнього і окремих значень один від одного, тобто розміру варіації параметра, служать дисперсія  $D$  і стандарт (середньоквадратичне відхилення)  $St$ , які є розмірними величинами і визначаються як

$$D = St^2 = \sum_{i=1}^N (x_i - \bar{x})^2 / N.$$

При зіставленні дисперсій кількох груп, що мають різні одиниці виміру, доцільно користуватися коефіцієнтом варіації  $V$

$$V = \frac{St}{\bar{x}} \cdot 100\%.$$

Для з'ясування особливостей розподілу параметрів вибірки можна також користуватися способом гістограм (варіаційних кривих). Гістограма являє собою графік розподілу числа випадків (частот) в різних інтервалах зміни досліджуваного параметра (рис. 1.1, 1.2). Для побудови гістограми весь діапазон  $L$  значень  $x$  розбивають на розряди й підраховують кількість значень  $\Delta N$ , що припадають на кожен  $k$ -тий розряд.  $\Delta N$  потім ділять на загальну кількість спостережень  $N$ . По осі абсцис відкладають розряди, а по осі ординат – відповідні цим розрядам частоти  $N_k$  (%).

При побудові гістограми не існує строгих методів вибору кількості розрядів  $K$ . Зазвичай користуються такими емпіричними правилами: визначають  $K = \sqrt{N}$  або  $K$  знаходять по інтервалу групування  $\Delta x$  вихідних даних, рівному згідно з формулою Стердесса

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3.32 \lg N}.$$

При цьому в кожному розряді має бути, як правило, не менше чотирьох-п'яти спостережень, а загальне число розрядів – не менше п'яти-шести. Таким чином, загальне число вимірювань повинно бути не менше двадцяти.

Для отримання варіаційної кривої значення  $N_k$  приписують середині інтервалу. Отримані точки з'єднують ламаною лінією.

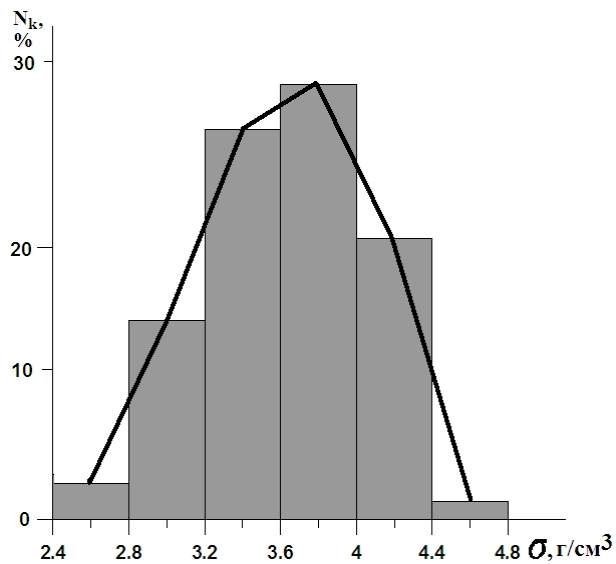


Рис. 1.1. Гістограма розподілу щільності

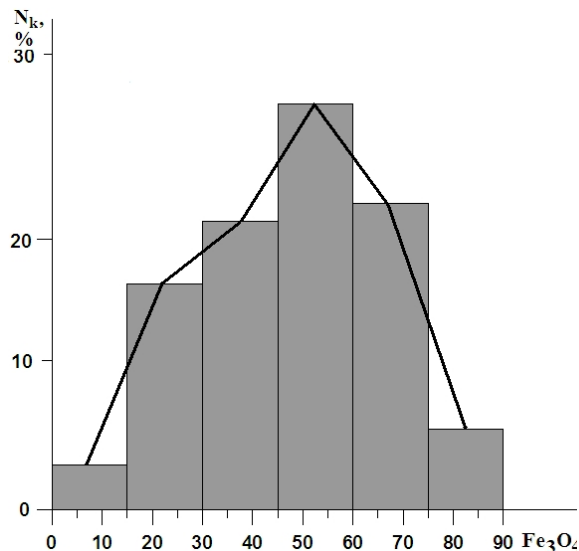


Рис. 1.2. Гістограма розподілу вмісту магнетиту

Середнє арифметичне значення  $\bar{x}$  досліджуваного параметра знаходиться як абсциса вертикальної лінії, що ділить площу всієї варіаційної кривої, на дві рівні частини. У разі симетричного характеру кривої розподілу значення  $\bar{x}$  відповідає абсцисі максимуму.

Варіаційна крива може бути використана для визначення закону розподілу досліджуваного параметра. Найбільш поширеним законом розподілу випадкової величини в природі є нормальний закон розподілу або закон похибок, який описується функцією:

$$f(x) = \frac{1}{St\sqrt{2\pi}} \cdot e^{-\frac{(x_i - \bar{x})^2}{St^2}},$$

де  $x_i$  – значення параметра;  $\bar{x}$  – середнє арифметичне параметра;  $St$  – стандарт розподілу параметра.

Основна особливість нормального закону в тому, що випадкова величина є сумою великого числа малих незалежних випадкових величин. Цей закон означає, що: похибка вимірів (або відхилення значень параметра від середнього) є неперервна випадкова величина; похибки однакової величини і рівних знаків зустрічаються однаково часто; великі похибки зустрічаються рідше, ніж малі.

Для перевірки, наскільки вимірні значення задовольняють нормальному закону розподілу, можна використати побудову кривих  $P(x)$  накопичення частот  $N_K$  в спеціальному «вірогіднісному» масштабі, який запропонований А.Г. Тарховим. Для обчислення величини  $P(x)$  підсумовується кількість випадків зі значеннями параметра, меншими або рівними  $x$ . «Вірогіднісний» масштаб складається для осі ординат  $P(x)$  таким чином, що в разі нормального розподілу точки  $P(x)$  розташуються на прямій лінії. Середнє арифметичне значення параметра визначається як абсциса точки перетину графіка  $P(x)$  з ординатою 50% (рис. 1.3).

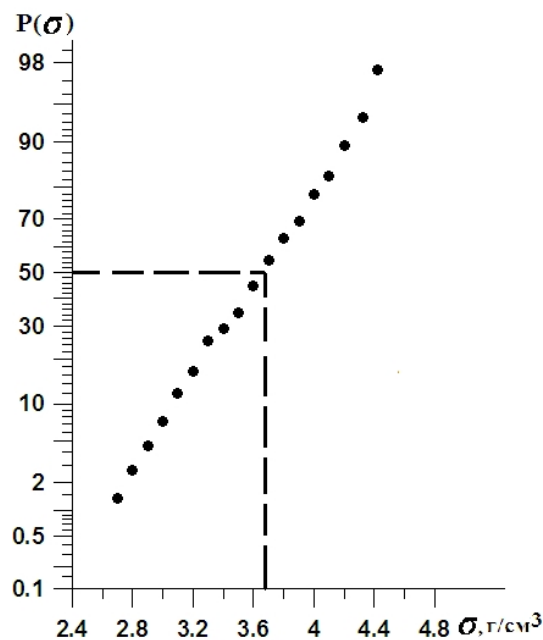


Рис. 1.3. Графік  $P(x)$  накопичення частот  $N_K$  в «вірогіднісному» масштабі

**Кореляційний аналіз статистичних зв'язків.** Важливою областю в дослідженні петрофізичних характеристик є вивчення зв'язку різних фізичних параметрів між собою або з іншими, що характеризують породи, величинами (особливостями хімічного і мінералогічного складів і т.д.). Якби кожне значення одного параметра однозначно визначало значення іншого, то можна було б говорити про наявність функціонального зв'язку між величинами, що розглядаються. Насправді ж доводиться мати справу зі статистичними зв'язками, при яких кожному значенню однієї величини відповідає статистичний ряд значень іншої величини. Математична статистика надає засоби виявлення і вивчення статистичних зв'язків між окремими параметрами. Обмежимося найбільш простим випадком – вивченням лінійного зв'язку між двома параметрами.

Більшість нелінійних залежностей можна привести до лінійних способами випрямлення. Наприклад, парабола другого порядку  $y = ax^2 + bx + c$  приводиться до прямої, якщо замість  $y$  розглядати  $z = (y - y_1)/(x - x_1)$  (тут  $x_1$  і

$y_1$  – координати довільної точки на кривій). Більш ретельно засоби отримання нелінійних залежностей розглянуто у наступній лабораторній роботі.

Знайти лінійний зв'язок між двома величинами  $x$  і  $y$  – означає знайти коефіцієнти рівняння:

$$y^* = ax + b, \quad (1.1)$$

яке носить назву рівняння регресії  $y$  на  $x$ . Рішення цього завдання починається з систематизації вихідних даних, які подаються у вигляді графіка, який названо полем кореляції. На графіку результати вивчення параметрів (щільності в  $\text{г/см}^3$  і вмісту магнетиту в % у зразках залізних руд) представляються точками в прямокутній системі координат (рис. 1.4). Проведена на рисунку лінія – лінія регресії. Її проводять за правилом найменших квадратів, тобто сума квадратів відхилень реальних значень від значень на лінії регресії повинна бути найменшою:

$$\sum_{i=1}^N [y_i - y^*(x_i)]^2 = \min.$$

Якщо записати  $y^*(x)$  як функцію від аргументу невідомих коефіцієнтів  $a$  і  $b$ , то для знаходження коефіцієнтів слід виконати диференціювання виразу (1.1) за цими коефіцієнтами, а похідні прирівняти нулю. Тоді отримуємо систему двох нормальних рівнянь:

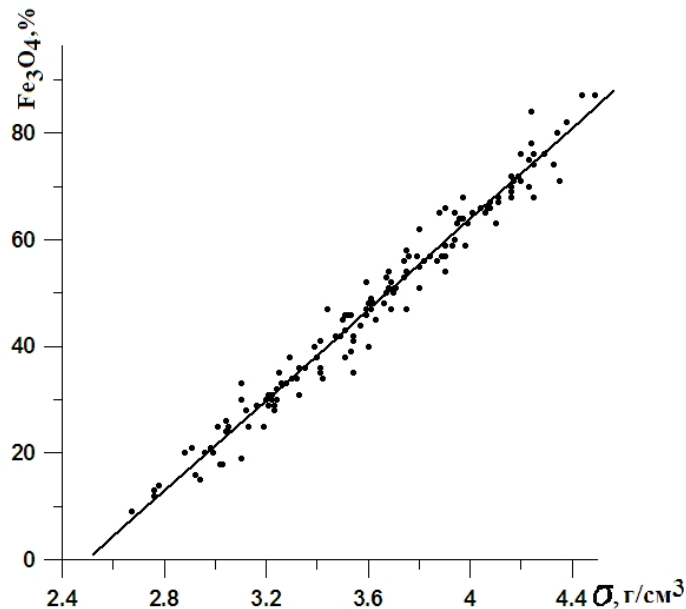


Рис. 1.4. Поле кореляції вмісту магнетиту і щільності

$$\begin{cases} [y_i - (ax_i + b)]x_i = 0 \\ [y_i - (ax_i + b)] = 0 \end{cases},$$

З рішення цієї системи одержуємо:

$$a = \frac{N \cdot \sum (x_i y_i) - \sum (x_i) \sum y_i}{N \cdot \sum x_i^2 - (\sum x_i)^2};$$

$$b = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum (x_i y_i)}{N \cdot \sum x_i^2 - (\sum x_i)^2}$$

Рівняння регресії може бути записано і в такому вигляді:

$$y - \bar{y} = r \cdot \frac{St_y}{St_x} \cdot (x - \bar{x}),$$

де  $\bar{x}$  і  $\bar{y}$  – середні значення  $x$  і  $y$  в заданій сукупності;  $St_x$  и  $St_y$  – середньоквадратичні відхилення (стандарти)  $x$  і  $y$  від середніх значень;  $r$  – коефіцієнт кореляції, який визначається за формулою:

$$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{St_x \cdot St_y}.$$

За коефіцієнтом кореляції, який змінюється в межах від -1 до +1, судять про силу статистичного зв'язку між  $x$  і  $y$ . При  $r = 0$  можна стверджувати, що лінійний зв'язок між  $x$  і  $y$  відсутній, хоча зв'язки іншого роду можуть мати місце. При  $r = \pm 1$  зв'язок функціональний. Тобто, чим більше  $|r|$ , тим ближче статистичний зв'язок між  $x$  і  $y$  до лінійної залежності.

Достовірність виявленого лінійного зв'язку (або, що те ж саме, достовірність знака  $r$  оцінюється за величиною надійності коефіцієнта кореляції  $|r|/\sigma_2$ , де  $\sigma_2 = (1 - r^2)/\sqrt{N}$  – похибка у визначенні  $r$ . Коефіцієнт кореляції має достовірний знак з гарантованою вірогідністю  $\varphi$ , якщо  $|r|/\sigma_2$  більше деякої величини  $t$ , залежної від  $\varphi$  і  $N$ . При великій кількості точок ( $N \geq 50$ ) зв'язок достовірний (з вірогідністю 0,999), якщо  $|r|/\sigma_2 \geq 3$ . При менших  $N$  для оцінки достовірності виявленого зв'язку слід користуватися спеціальними таблицями  $t(\varphi, N)$ .

Для вирішення різних математичних і геологічних задач, в тому числі і для статистичної обробки вимірювань широко використовуються комп'ютери, що дозволяє швидко і досить точно обробляти великі масиви числового матеріалу.

### 3. Завдання та порядок виконання роботи

1. Побудувати гістограми і варіаційні криві розподілів щільності і вмісту магнетиту.

2. Побудувати поле кореляції вмісту магнетиту і щільності.

3. Розрахувати середньоарифметичні, дисперсії, стандарти, коефіцієнти варіації, коефіцієнт кореляції, надійність і коефіцієнти регресії.

Примітка. Кожен студент виконує один з варіантів, отриманий у викладача.

#### **4. Вимоги до звіту**

У звіті необхідно привести вихідні дані, виконати необхідні графічні побудови та розрахунки (з докладними поясненнями!) згідно з завданням.

#### **5. Контрольні питання**

1. Які основні процедури складають первинну статистичну обробку геологічної інформації?

2. Перерахувати основні статистичні характеристики випадкової величини.

3. Що характеризує середнє арифметичне сукупності?

4. Що характеризує дисперсія сукупності?

5. Що характеризує гістограма розподілу сукупності?

6. В чому полягає основна особливість нормального закону розподілу?

7. Що означає термін «статистичний зв'язок» між параметрами?

8. Поясніть суть методу найменших квадратів.

9. За допомогою чого і яким чином судять про силу статистичного зв'язку між параметрами?

### **Лабораторна робота №2**

#### **Визначення виду аналітичної залежності**

##### **1. Мета роботи**

Вивчення методів і прийомів визначення виду аналітичної залежності при знаходженні зв'язку між емпіричними даними.

##### **2. Теоретичні основи**

Аналіз різних процесів призводить до необхідності виявлення істотних факторів, що впливають на досліджуваний процес, а також до вибору форми зв'язку між факторами і до оцінки параметрів отриманих рівнянь.

Будемо вважати, що деяке явище характеризується двома змінними величинами  $x$  і  $y$ , з яких  $x$  вибирається як незалежна, а  $y$  – як залежна змінна

величина. Припустимо, що між змінними  $x$  і  $y$  існує однозначна відповідність, тобто кожному значенню незалежної змінної величини  $x$  відповідає із заданим ступенем точності одне значення залежної змінної  $y$ . У цьому випадку виникає задача виявлення форми зв'язку і визначення формульної залежності, яка задає  $y$  як функцію  $f(x)$ .

У загальному вигляді завдання можна сформулювати наступним чином. Нехай у результаті досліджень деякої величини  $x$  зі значеннями  $x_1, x_2, \dots, x_n$  поставлені у відповідність значення  $y_1, y_2, \dots, y_n$  деякої величини  $y$ . Потрібно підібрати вид аналітичної залежності  $y = f(x)$ , що зв'язує змінні  $x$  і  $y$ .

Будемо називати аналітичні залежності, що отримано в результаті спостережень, *емпіричними*. Виявлення емпіричних залежностей ділиться на два основних етапи – *вибір емпіричної формули та уточнення коефіцієнтів обраної формули*.

## 2.1 Вибір емпіричних формул для нелінійних залежностей

На практиці зазвичай намагаються представити аналітичну залежність  $y$  вигляді прямої, яка задається формулою  $y = ax + b$ . Однак, це не завжди вдається. Графічна побудова нелінійної залежності не дає відповіді на питання про те, який аналітичний вигляд має ця функція, тобто чи буде ця залежність ступеневою, дрібно-раціональною, логарифмічною і т.д.

Нехай  $y$  являє собою функцію однієї змінної з двома параметрами  $a$  і  $b$ . В якості набору функцій, з яких будемо вибирати емпіричну залежність, розглянемо:

- 1) лінійну функцію  $y = ax + b$ ;
- 2) показникову функцію  $y = ab^x$ ;
- 3) дрібно-раціональну функцію виду  $y = 1/(ax + b)$ ;
- 4) логарифмічну функцію  $y = a \ln x + b$ ;
- 5) ступеневу функцію  $y = ax^b$  (вона визначає параболічну залежність, якщо параметр  $b > 0$ , і гіперболічну залежність, якщо  $b < 0$ ; якщо ж параметр  $b = 0$ , то залежність вироджується в лінійну);
- 6) гіперболічну функцію  $y = a + b/x$ ;
- 7) дрібно-раціональну функцію виду  $y = x/(ax + b)$ .

Для найкращого вибору виду аналітичної залежності  $y = f(x, a, b)$ , що відповідає побудованому графіку, виконаємо наступні проміжні обчислення. На заданому відрізку зміни незалежної змінної  $x$  виберемо точки, досить надійні і, по можливості, далеко віддалені одна від одної. Для простоти будемо вважати, що це точки  $x_1$  і  $x_n$ . Обчислимо середнє арифметичне, середнє геометричне і середнє гармонійне



$$x_{ар} = \frac{x_1 + x_n}{2}, \quad x_{геом} = \sqrt{x_1 x_n}, \quad x_{гарм} = \frac{2x_1 x_n}{x_1 + x_n}.$$

За обчисленими значеннями незалежної змінної знайдемо з побудованого графіка відповідні значення залежної змінної

$$x_{ар} \rightarrow y_1^*, \quad x_{геом} \rightarrow y_2^*, \quad x_{гарм} \rightarrow y_3^*.$$

для поки ще невідомої аналітичної залежності  $y = f(x, a, b)$ . Виконаємо допоміжні обчислення для залежної змінної. Обчислимо середнє арифметичне, середнє геометричне і середнє гармонійне крайніх значень

$$y_{ар} = \frac{y_1 + y_n}{2}, \quad y_{геом} = \sqrt{y_1 y_n}, \quad y_{гарм} = \frac{2y_1 y_n}{y_1 + y_n}.$$

Порівняємо знайдені з графіка  $y_1^*, y_2^*, y_3^*$  з обчисленими  $y_{ар}, y_{геом}, y_{гарм}$  і оцінимо такі похибки:

$$\begin{aligned} |y_1^* - y_{ар}| &= \varepsilon_1, & |y_1^* - y_{геом}| &= \varepsilon_2, & |y_1^* - y_{гарм}| &= \varepsilon_3, \\ |y_2^* - y_{ар}| &= \varepsilon_4, & |y_2^* - y_{геом}| &= \varepsilon_5, \\ |y_3^* - y_{ар}| &= \varepsilon_6, & |y_3^* - y_{гарм}| &= \varepsilon_7. \end{aligned}$$

Знайдемо з цих похибок мінімальну:  $\varepsilon = \min(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_7)$ .

1 Якщо найменшою серед усіх абсолютних похибок виявиться  $\varepsilon_1$ , то в якості аналітичної залежності для даного графіка хорошим наближенням служить лінійна функція  $y = ax + b$ .

2 Якщо найменшою абсолютною похибкою є  $\varepsilon_2$ , то в якості емпіричної залежності слід вибрати показникову функцію  $y = ab^x$ .

3 У тому випадку, коли найменша з абсолютних похибок є  $\varepsilon_3$ , шукана емпірична залежність визначається дрібно-раціональною функцією виду  $y = 1/(ax + b)$ .

4 Якщо найменша з абсолютних похибок є  $\varepsilon_4$ , то хорошим наближенням служить логарифмічна функція  $y = a \ln x + b$ .

5 Для випадку, коли найменшою абсолютною похибкою є  $\varepsilon_5$ , в якості емпіричної залежності вибирається ступенева функція  $y = ax^b$ .

6 Якщо найменшою з абсолютних похибок виявиться  $\varepsilon_6$ , то за шукану залежність слід вибрати гіперболічну  $y = a + b/x$ .

7 Нарешті, в тому випадку, коли найменша з абсолютних похибок є  $\varepsilon_7$ , в якості аналітичної залежності слід вибирати дрібно-раціональну функція виду  $y = x/(ax + b)$ .

**Приклад 1.** Підібрати емпіричну залежність для функції заданої таблично:

x	1	2	3	4	5	6	7	8	9
y	521	308	240,5	204	183	171	159	152	147

*Рішення.* 1) Припустимо, що в даному прикладі крайні табличні значення досить надійні. Проведемо допоміжні обчислення і знайдемо для крайніх значень середнє арифметичне  $x_{ар} = 5$ , середнє геометричне  $x_{геом} = 3$  і середнє гармонійне  $x_{гарм} = 1,8$ .

2) З графіка (рис. 2.1) знайдемо значення функції, що відповідають обчисленим значенням аргументу: для  $x_{ар} = 5$  маємо  $y_1^* \approx 180$ ; для  $x_{геом} = 3$  маємо  $y_2^* \approx 240$  для  $x_{гарм} = 1,8$  маємо  $y_3^* \approx 341$ .

3) Виконаємо додаткові розрахунки для залежної змінної. Знайдемо для крайніх значень середнє арифметичне  $y_{ар} = 334$ , середнє геометричне  $y_{геом} = 274$ , середнє гармонічне  $y_{гарм} = 228$ .

4) Порівняємо знайдені графічно значення залежної змінної  $y_1^*$ ,  $y_2^*$ ,  $y_3^*$  з  $y_{ар}$ ,  $y_{геом}$ ,  $y_{гарм}$ .

$$\begin{aligned} \varepsilon_1 &= |180 - 334| = 154, & \varepsilon_2 &= |180 - 274| = 106, & \varepsilon_3 &= |180 - 228| = 48, \\ \varepsilon_4 &= |240 - 334| = 94, & \varepsilon_5 &= |240 - 274| = 34, & \varepsilon_6 &= |341 - 334| = 7, \\ \varepsilon_7 &= |341 - 228| = 113. \end{aligned}$$

Оскільки найменша з абсолютних похибок є  $\varepsilon_6$ , в якості аналітичної залежності слід вибрати гіперболічну залежність  $y = a + b/x$ .

## 2.2 Уточнення коефіцієнтів

Для уточнення коефіцієнтів обраної аналітичної залежності  $y = f(x, a, b)$  можна скористатися кількома методами.

**Метод обраних точок.** На побудованій кривій візьмемо дві довільні точки  $M(x_1^*; y_1^*)$  і  $N(x_2^*; y_2^*)$ , складемо систему

$$\begin{cases} y_1^* = f(x_1^*, a, b), \\ y_2^* = f(x_2^*, a, b), \end{cases}$$

і вирішимо її відносно шуканих параметрів  $a$  і  $b$ .

**Метод середніх.** У емпіричну формулу  $y = f(x, a, b)$  послідовно підставляємо табличні значення  $x_i (i = 1, 2, \dots, n)$ . Отримані значення функції  $y_i = f(x_i, a, b)$ , будуть відхилятися від табличних значень:  $y_i - f(x_i, a, b) = \varepsilon_i$ . Згідно з методом середніх, за найкраще положення кривої приймається те, для якого сума алгебраїчних відхилень дорівнює нулю, тобто

$$\sum_{i=1}^n \varepsilon_i = \sum_{i=1}^n [y_i - f(x_i, a, b)] = 0.$$

Для визначення параметрів  $a$  і  $b$  за методом середніх вся сукупність  $\varepsilon_i (i = 1, 2, \dots, n)$  розбивається на дві групи так, щоб алгебраїчна сума відхилень кожної групи дорівнювала нулю. Таким чином, для визначення параметрів  $a$  і  $b$  отримуємо наступну систему двох рівнянь:

$$\begin{cases} \sum_{j=1}^l [y_j^I - f(x_j^I, a, b)] = 0, \\ \sum_{k=1}^{n-l} [y_k^{II} - f(x_k^{II}, a, b)] = 0, \end{cases}$$

де  $l$  і  $n-l$  – кількість табличних даних відповідно для першої та другої групи. Замінивши суму різниць різницею сум в кожному рівнянні системи, отримаємо

$$\begin{cases} \sum_{j=1}^l y_j^I = \sum_{j=1}^l f(x_j^I, a, b), \\ \sum_{k=1}^{n-l} y_k^{II} = \sum_{k=1}^{n-l} f(x_k^{II}, a, b). \end{cases}$$

Спільне рішення цієї системи визначає чисельне значення двох параметрів  $a$  і  $b$ , підставивши які в  $y = f(x, a, b)$ , отримаємо шукане емпіричне співвідношення.

**Метод найменших квадратів.** Згідно з методом найменших квадратів, найкращими параметрами  $a$  і  $b$  вважаються ті, для яких сума квадратів відхилень мінімальна:

$$F(a, b) = \sum_{i=1}^n [y_i - f(x_i, a, b)]^2 = \min.$$

Знайдемо частинні похідні функції  $f(a, b)$  по варійованим параметрам  $a$  і  $b$ :

$$\frac{\partial F(a, b)}{\partial a} = -2 \sum_{i=1}^n [y_i - f(x_i, a, b)] f'_a(x_i, a, b),$$

$$\frac{\partial F(a, b)}{\partial b} = -2 \sum_{i=1}^n [y_i - f(x_i, a, b)] f'_b(x_i, a, b).$$

В силу необхідної умови екстремуму функції багатьох змінних, найкращими значеннями параметрів  $a$  і  $b$  служать ті, при яких частинні похідні цієї функції по варійованим параметрам обертаються в нуль:

$$\begin{cases} \frac{\partial F(a, b)}{\partial a} = 0, \\ \frac{\partial F(a, b)}{\partial b} = 0, \end{cases}$$

тобто

$$\begin{cases} \sum_{i=1}^n [y_i - f(x_i, a, b)] f'_a(x_i, a, b) = 0, \\ \sum_{i=1}^n [y_i - f(x_i, a, b)] f'_b(x_i, a, b) = 0. \end{cases}$$

Рішення цієї системи щодо  $a$  і  $b$  дає шукане найкраще значення числових параметрів.

**Метод перетворення координат.** Розглянемо в системі координат  $xOy$  деяку нелінійну залежність  $y = f(x, a, b)$ , безперервну і монотонну на замкнутому відрізку  $[x_1, x_n]$ . Перейдемо до нових змінних  $q = \varphi(x)$  і  $z = \psi(y)$  так, щоб в новій системі координат  $qOz$  емпірична залежність стала лінійною:  $z = aq + b$ .

Точки  $N_i$  з координатами  $[\varphi(x_i); \psi(y_i)]$  в площині  $qOz$  практично лежать на прямій лінії. Або, якщо при побудові на площині  $qOz$  виявиться, що точки лежать на прямій лінії, то між змінними  $q$  і  $z$  має місце залежність  $\psi(y) = a\varphi(x) + b$ .

Деякі нелінійні залежності перетворенням координат можна звести до лінійної наступними прийомами.

1) Для показникової залежності виду  $y = ab^x$ . Логарифмуючи, маємо  $\lg y = x \lg b + \lg a$ . Прийmemo  $\lg b = B$ ,  $\lg a = A$ ,  $\lg y = z$ ,  $x = q$  і в площині  $qOz$  одержимо рівняння прямої  $z = Bq + A$ .

2) Дробово-лінійну залежність  $y = 1/(ax + b)$  також можна перетворити в лінійну. Дійсно, знайдемо зворотню залежність  $1/y = (ax + b)$  і введемо нові змінні  $q = x$ ,  $z = 1/y$ ; тоді отримаємо залежність виду  $z = aq + b$ .

3) Для логарифмічної залежності  $y = a \ln x + b$  введемо нові змінні  $q = \lg x$ ,  $z = y$ . Тоді знову отримаємо лінійну залежність  $z = aq + b$ .

4) Розглянемо ступеневу залежність  $y = ax^b$ , де  $a > 0$ ,  $b > 0$ . Логарифмуючи, знаходимо  $\lg y = a \lg x + \lg b$ , звідки, вважаючи  $z = \lg y$ ,  $q = \lg x$ ,  $B = \lg b$ , маємо  $z = aq + B$ . Пряму отриманого виду зручно будувати, якщо осі  $q$  і  $z$  у площині  $qOz$  взяти в логарифмічному масштабі. Початком координат логарифмічною сітки служить точка  $q = 0$ ,  $z = B$ .

5) У виразі  $y = a + b/x$  зробимо заміну змінних  $1/x = q$ ,  $y = z$ . Тим самим задана гіперболічна залежність перетвориться в лінійну  $z = aq + b$ .

6) Розглянемо дрібно-раціональну функцію  $y = x/(ax + b)$ . Функція, зворотна даної, має вигляд  $1/y = a + b/x$ . Ввівши нові координати  $z = 1/y$ ,  $q = 1/x$ , знову отримаємо лінійну залежність  $z = aq + b$ .

**Приклад 2.** Дослідні дані занесені в наступну таблицю:

$x$	1	2	3	4	5	6	7	8	9
$y$	0,33	0,49	0,59	0,65	0,71	0,755	0,77	0,81	0,82

Скласти емпіричну формулу.

*Рішення.* Побудуємо графік (рис. 2.2) і визначимо тип кривої так, як це було зазначено в пункті 2.1. Аналітична функція, що досить добре відповідає табличним даним являє собою дрібно-раціональну функцію виду  $y = x/(ax + b)$ .

Розглянемо функцію  $1/y = a + b/x$ , зворотню даній, і введемо нові змінні  $z = 1/y$ ,  $q = 1/x$ . Тоді в площині  $qOz$  отримаємо лінійну залежність  $z = aq + b$ . Для того, щоб знайти коефіцієнти  $q$  і  $z$ , складемо таблицю:

$q$	1	0,5	0,333	0,25	0,2	0,166	0,143	0,125	0,111
$z$	3	2	1,66	1,54	1,41	1,32	1,3	1,24	1,22

Побудуємо в площині  $qOz$  точки, що відповідають табличним значенням (рис. 2.3). Ми бачимо, що ці точки лежать на одній прямій. Отже, наші припущення про тип кривої виявилися правильними. Залишається визначити параметри  $a$  і  $b$ :  $a = 1$ ;  $b = (3 - 1)/(1 - 0) = 2$ . Таким чином, емпірична формула має вигляд  $y = x/ + 2$ .

### 3. Порядок виконання роботи

1 Винести на графік дискретні значення  $y = f(x)$ , відповідно до отриманого варіанту завдання.

2 Побудувати графік функції  $y = f(x)$ , яка повинна бути монотонною і однозначною.

3 Визначити тип залежності (див. пункт. 2.1 і приклад 1)

4 Визначити коефіцієнти отриманої залежності за способом, вказаним викладачем (див. пункт 2.2 і приклад 2).

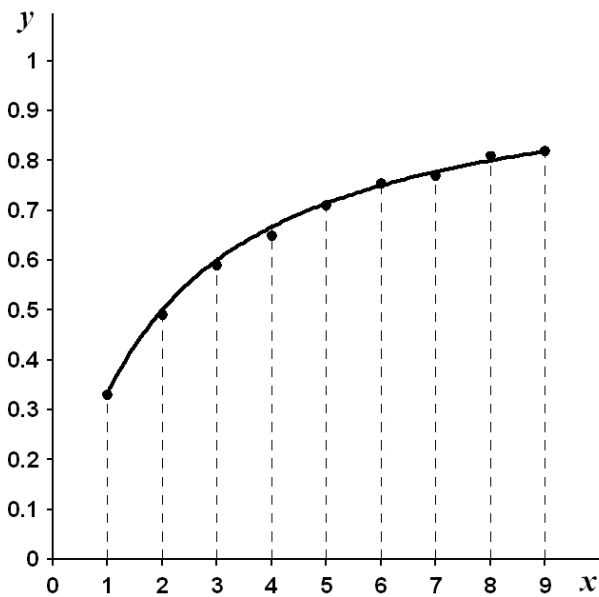


Рис. 2.2. Графік нелінійної функції  $f(x)$

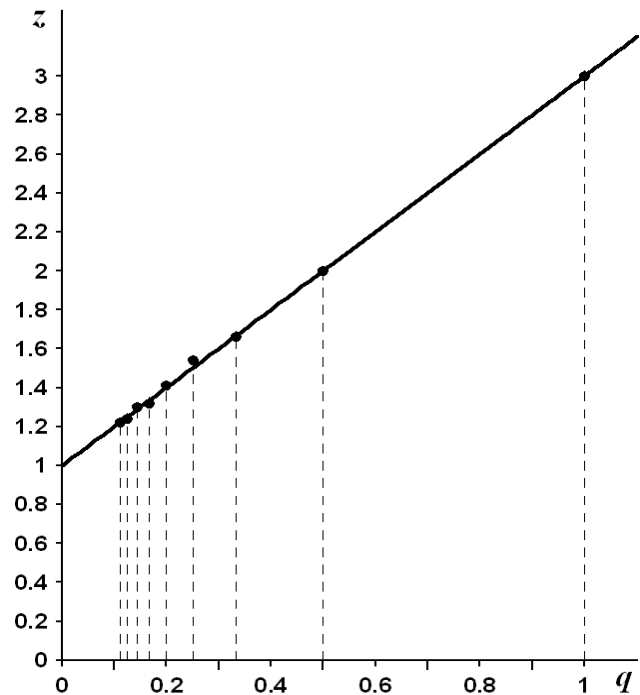


Рис. 2.3. Графік функції  $f(x)$  після перетворення координат

#### 4. Вимоги до звіту

У звіті необхідно привести вихідні дані, виконати необхідні графічні побудови та розрахунки (з докладними поясненнями!). В якості основи слід орієнтуватися на приклади 1 і 2.

#### 5. Контрольні питання

1. В чому полягає необхідність визначення виду аналітичної залежності, відмінної від лінійної?

2. Як обчислюються середнє арифметичне, середнє геометричне і середнє гармонійне?

3. Пояснити, як уточнюють коефіцієнти нелінійної залежності методом обраних точок?

4. Пояснити, як уточнюють коефіцієнти нелінійної залежності методом середніх?

5. Пояснити, як уточнюють коефіцієнти нелінійної залежності методом найменших квадратів?

6. Пояснити, як уточнюють коефіцієнти нелінійної залежності методом перетворення координат?

## **Лабораторна робота №3** **Тренд-аналіз геологічних даних**

### **1. Мета роботи**

Освоєння методики аналізу полів геологічних параметрів, поділ їх на складові різних рівнів (порядків) за допомогою тренд-аналізу.

### **2. Теоретичні основи**

Графічне зображення просторових змін геологічних параметрів у вигляді графіків, розрізів, профілів, карт широко поширене в геологічній практиці і стало традиційним. На цих геологічних документах зазвичай виділяють як напрям зміни (зростання або зменшення) досліджуваної ознаки, так і положення аномальних зон (в розрізі або на площі). Особливої уваги заслуговує проблема виділення регіональних напрямків зміни геологічного параметра – напрямків тренду. Так, наприклад, напрямок зміни гранулометричного складу вказує на положення області денудації – джерела знесення; напрямок регіонального збільшення продуктивності нафтоносних структур може бути пов'язано з положенням області генерації вуглеводнів і т. д. В умовах порівняно простої геологічної будови (або слабкої вивченості) такі регіональні напрямки досить впевнено виділяються на відповідних картах. Однак у більш складних умовах при мозаїчному характері розподілу локальних аномалій досліджуваного геологічного параметру виділення напрямків регіональної тенденції його зміни часто представляє складну задачу, в рішення якої вносяться суб'єктивні уявлення апріорних геологічних концепцій.

У загальному вигляді процедура тренд-аналізу може бути представлена у вигляді:

$$G(x, y) = T(x, y) + g(x, y), \quad (3.1)$$

де  $G(x, y)$  – функція параметру, що вивчається;  $T(x, y)$  – тренд (тенденція) параметра;  $g(x, y)$  – відхилення параметра від його тренда.

При цьому операція розкладання (3.1) виконується таким чином, що тренд  $T(x, y)$  є регулярною складовою функції  $G(x, y)$ , а відхилення  $g(x, y)$  має стохастичний характер. Оскільки тренд  $T(x, y)$  має регулярний характер, він може бути описаний рівнянням детермінованої функції. Тому для опису тренда часто використовують ступеневі поліноми різного порядку.

Практично зазвичай здійснюють послідовний підбір полінома все більш високого порядку, поступово наближаючи  $T(x, y)$  до  $G(x, y)$ . Природно, що зі збільшенням ступеня полінома  $T(x, y) \rightarrow G(x, y)$ , а  $g(x, y) \rightarrow 0$ .

У разі тренда першого порядку для апроксимації досліджуваної поверхні використовується рівняння площини:

$$T_1(x, y) = ax + by + c, \quad (3.2)$$

де  $a, b, c$  – сталі коефіцієнти.

Значення коефіцієнтів знайдемо за методом найменших квадратів з умови:

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0, \quad \frac{\partial S}{\partial c} = 0, \quad (3.3)$$

де

$$S = \sum_{i=1}^N [G(x_i, y_i) - ax_i - by_i - c_i]^2. \quad (3.4)$$

З умови (3.3) знайдемо систему лінійних рівнянь:

$$\begin{aligned} a \sum_{i=1}^N x_i^2 + b \sum_{i=1}^N x_i y_i + c \sum_{i=1}^N x_i &= \sum_{i=1}^N G(x_i, y_i) x_i; \\ a \sum_{i=1}^N x_i y_i + b \sum_{i=1}^N y_i^2 + c \sum_{i=1}^N y_i &= \sum_{i=1}^N G(x_i, y_i) y_i; \\ a \sum_{i=1}^N x_i + b \sum_{i=1}^N y_i + cN &= \sum_{i=1}^N G(x_i, y_i). \end{aligned} \quad (3.5)$$

З рішення системи (3.5) отримаємо значення сталих  $a, b, c$  і побудуємо в координатах  $xOy$  площину (3.2), орієнтація якої і характеризує найбільш загальну тенденцію зміни параметра  $G(x, y)$ .

Загалом поточне значення  $T(x, y)$  відображає деяке усереднене значення поля  $G(x, y)$ , що змінюється за площею. В цьому зв'язку особливий інтерес набуває визначення положення відхилень



$$g(x,y) = G(x,y) - T(x,y), \quad (3.6)$$

що утворюють на площі систему відносних позитивних і негативних аномалій поля. Справді, виявлення положення зон з відносно підвищеними значеннями, наприклад, пористості, може служити підставою для закладення нових свердловин.

Ступінь наближення поверхні тренда до спостережених результатів оцінюється за допомогою дисперсійного аналізу.

Дисперсія спостережених значень досліджуваного геологічного параметра розраховується за формулою:

$$\sigma_{cn}^2 = \frac{1}{N} \sum_{i=1}^N (G_i(x,y) - \bar{G})^2. \quad (3.7)$$

Дисперсія обчислених за рівнянням регресії значень досліджуваного геологічного параметра, розраховується за формулою:

$$\sigma_{розр}^2 = \frac{1}{N} \sum_{i=1}^N (T_i(x,y) - \bar{T})^2. \quad (3.8)$$

Однак більш наочними параметрами є стандарти або середні квадратичні відхилення від середніх значень досліджуваного геологічного параметра:

$$\sigma_{cn} = \sqrt{\sigma_{cn}^2}, \quad \sigma_{розр} = \sqrt{\sigma_{розр}^2}. \quad (3.9)$$

Різниця між цими величинами буде характеризувати мінливість відхилень від поверхні тренда (величини середньоквадратичних відхилень фактичних значень досліджуваного геологічного параметра від площин їх трендів):

$$\delta = \sigma_{cn} - \sigma_{розр}. \quad (3.10)$$

Відношення сум квадратів відхилень фактичних і розрахункових значень досліджуваного геологічного параметра від їх математичного очікування

$$R^2 = \frac{\sigma_{розр}^2}{\sigma_{cn}^2} \quad (3.11)$$

служить мірою мінливості досліджуваного геологічного параметра щодо площин їх трендів ( $R^2$  – коефіцієнт дискримінації).

Коефіцієнти множинної кореляції зв'язку розраховуються за формулою:

$$R = \sqrt{R^2} . \quad (3.12)$$

В якості вихідних графічних моделей полів геологічних параметрів використовуються структурні карти, геологічні розрізи, карти потужностей, колекторських властивостей порід, фізико-хімічних властивостей нафти та ін.

### 3. Порядок виконання роботи

Фактичні дані з карт або розрізів знімаються в вузлах квадратної сітки, що «покриває» вихідні графічні моделі. Параметри сітки визначаються розмірами локальних і зональних ускладнень, що простежуються на вихідних картах і розрізах.

Далі розраховують статистичні характеристики фактичних і трендових значень досліджуваного геологічного параметра:

- 1) сталі коефіцієнти  $a$ ,  $b$ ,  $c$  для (3.2) шляхом рішення системи лінійних рівнянь (3.5);
- 2) дисперсію спостережених значень по (3.7);
- 3) дисперсію обчислених за рівнянням регресії значень по (3.8);
- 4) стандарти, або середні квадратичні відхилення від середніх по (3.9);
- 5) мінливість відхилень фактичних значень від поверхні тренда по (3.10);
- 6) коефіцієнт дискримінації за (3.11);
- 7) коефіцієнт множинної кореляції по (3.12).

Далі слід побудувати на вихідній поверхні  $G(x, y)$  карту регіональних тенденцій (тренд)  $T(x, y)$  досліджуваного геологічного параметра.

Потім знаходимо за (3.6) у вузлах сітки величини відхилень фактичних значень досліджуваного геологічного параметра від трендових і будуємо карту локальних ускладнень (аномалій)  $g(x, y)$ .

Вихідні дані і результати обчислень представляють у вигляді таблиці 3.1.

Таблиця 3.1.

Результати обробки геологічних даних по тренд-аналізу I-го порядку

Номер точки	$X_i$	$Y_i$	$X_i^2$	$Y_i^2$	$X_i Y_i$	$G(x_i, y_i)$	$G(x_i, y_i) X_i$	$G(x_i, y_i) Y_i$	$T(x_i, y_i)$	$T^2(x_i, y_i)$	$g(x_i, y_i)$
1	2	3	4	5	6	7	8	9	10	11	12
	$\sum_{i=1}^N X_i$	$\sum_{i=1}^N Y_i$	$\sum_{i=1}^N X_i^2$	$\sum_{i=1}^N Y_i^2$	$\sum_{i=1}^N X_i Y_i$	$\sum_{i=1}^N G(x_i, y_i)$	$\sum_{i=1}^N G(x_i, y_i) X_i$	$\sum_{i=1}^N G(x_i, y_i) Y_i$	$\sum_{i=1}^N T(x_i, y_i)$	$\sum_{i=1}^N T^2(x_i, y_i)$	

#### *Приклад виконання роботи*

1. На рисунку 3.1 наведена карта ізоліній щільності шару пісковика. На ній відзначені точки («хрестики») зняття значень. Результати занесені в таблицю 3.2 (колонки 2, 3 і 7).

2. Обчислено суми, необхідні для складання системи рівнянь (3.5) – відповідні колонки і останній рядок в таблиці 3.2. Складена система рівнянь:

$$7371.00a + 3276.00b + 819.00c = 1511.36;$$

$$3276.00a + 2856.00b + 504.00c = 951.48;$$

$$819.00a + 504.00b + 126c = 237.20.$$

3. Рішення системи рівнянь<sup>1</sup> дало такі значення коефіцієнтів в рівнянні тренду:

$$a = -0.0149; \quad b = +0.0032; \quad c = +1.9660.$$

Рівняння площини тренда має вигляд:

$$T_1(x, y) = -0.0149x + 0.0032y + 1.9660.$$

4. З цього рівняння визначені значення тренда в тих же точках, в яких зняті вихідні дані значення (колонка 10 таблиці 3.2) і побудована карта тренда, поєднана з картою вихідних даних (рис. 3.2).

5. Знайдена різниця вихідних значень і тренда (колонка 12 таблиці 3.2). На рисунку 3.3 наведена карта локальної зміни щільності пісковика. Слід відзначити такі особливості локальних змін щільності пісковика: знижені значення спостерігаються в центральній частині зі зміщенням на північ, мінімум розташовується в районі точки з координатами  $x = 6$ ,  $y = 6$ . Знижені значення спостерігаються і в центрі південної границі ділянки. На захід локальні значення зростають швидше, ніж на схід.

6. Обчислено дисперсію спостережених значень по (3.7)  $\sigma_{cn}^2 = 0.0040$ ; дисперсія обчислених за рівнянням регресії значень по (3.8)  $\sigma_{розр} = 0.0037$ ; стандарти (середні квадратичні відхилення від середніх) по (3.9)  $\sigma_{cn} = 0.063$ ,  $\sigma_{розр} = 0.060$ ; мінливість відхилень фактичних значень від поверхні тренда по

<sup>1</sup> Вирішити систему з трьох лінійних рівнянь з трьома невідомими зручно методом визначників. З урахуванням

позначень в даній роботі система має вигляд: 
$$\begin{cases} \alpha_1 a + \beta_1 b + \gamma_1 c = \delta_1 \\ \alpha_2 a + \beta_2 b + \gamma_2 c = \delta_2 \\ \alpha_3 a + \beta_3 b + \gamma_3 c = \delta_3 \end{cases}$$
, где  $a, b, c$  – невідомі.  $a = \Delta_a / \Delta$ ;

$b = \Delta_b / \Delta$ ;  $c = \Delta_c / \Delta$ .

$\Delta = \alpha_1 \beta_2 \gamma_3 - \alpha_1 \beta_3 \gamma_2 + \beta_1 \gamma_2 \alpha_3 - \beta_1 \gamma_3 \alpha_2 + \gamma_1 \alpha_2 \beta_3 - \gamma_1 \alpha_3 \beta_2$

$\Delta_a = \delta_1 \beta_2 \gamma_3 - \delta_1 \beta_3 \gamma_2 + \beta_1 \gamma_2 \delta_3 - \beta_1 \gamma_3 \delta_2 + \gamma_1 \delta_2 \beta_3 - \gamma_1 \delta_3 \beta_2$

$\Delta_b = \alpha_1 \delta_2 \gamma_3 - \alpha_1 \delta_3 \gamma_2 + \delta_1 \gamma_2 \alpha_3 - \delta_1 \gamma_3 \alpha_2 + \gamma_1 \alpha_2 \delta_3 - \gamma_1 \alpha_3 \delta_2$

$\Delta_c = \alpha_1 \beta_2 \delta_3 - \alpha_1 \beta_3 \delta_2 + \beta_1 \delta_2 \alpha_3 - \beta_1 \delta_3 \alpha_2 + \delta_1 \alpha_2 \beta_3 - \delta_1 \alpha_3 \beta_2$

(3.10)  $\delta = 0.003$  ; коефіцієнт дискримінації за (3.11)  $R^2 = 0.913$ ; коефіцієнт множинної кореляції по (3.12)  $R = 0.95$ .

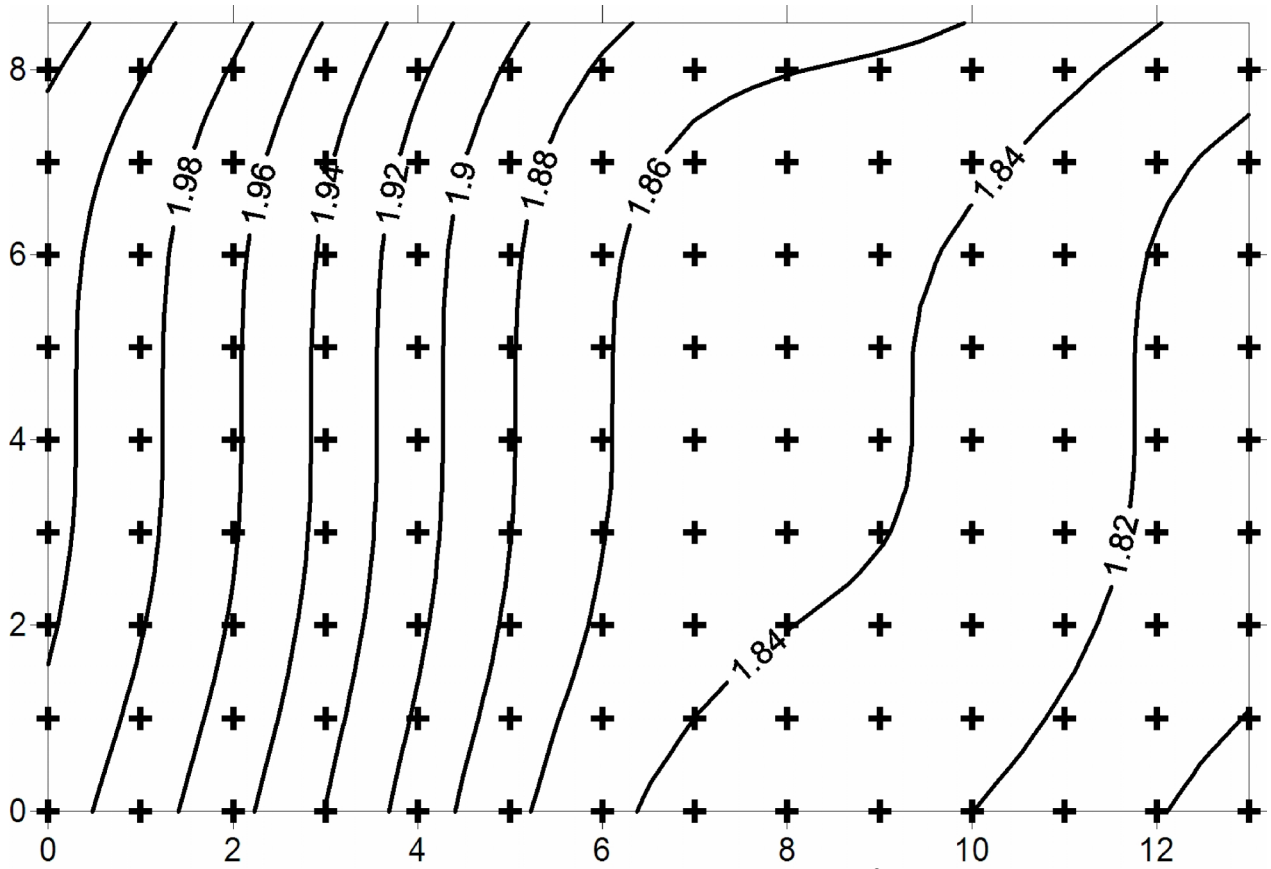


Рис. 3.1. Карта ізолій щільності (в  $\text{г/см}^3$ ) шару пісковика

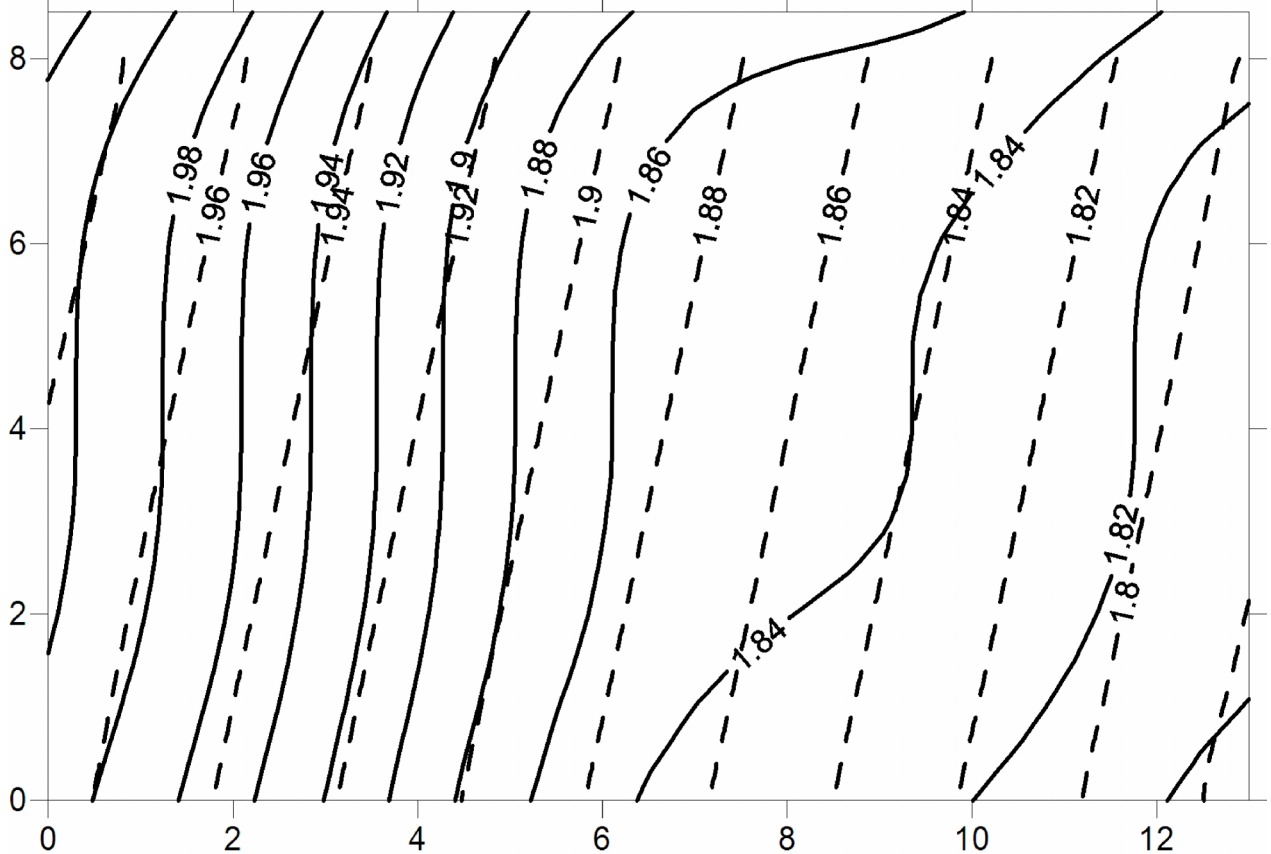


Рис. 3.2. Карта тренду (ізолінії значень щільності показані пунктиром), поєднана з картою вихідних даних

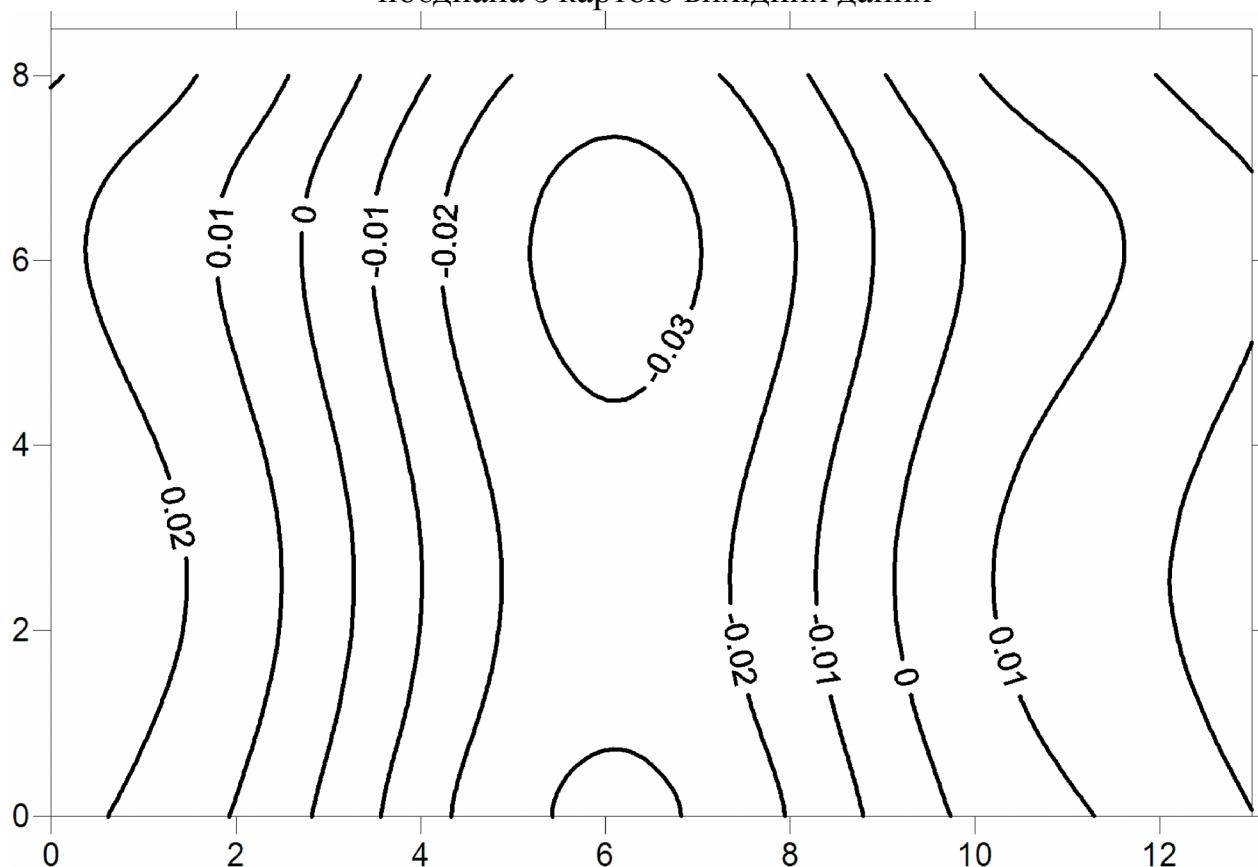


Рис. 3.3. Карта локального зміни щільності пісковика

Таблиця 3.2.

Результати обробки розподілу щільності пісковика за допомогою тренд-аналізу

Номер точки	$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$	$G(x_i, y_i)$	$G(x_i, y_i) x_i$	$G(x_i, y_i) y_i$	$T(x_i, y_i)$	$T^2(x_i, y_i)$	$g(x_i, y_i)$
1	0.00	0.00	0.00	0.00	0.00	1.99	0.00	0.00	1.97	3.87	0.02
2	1.00	0.00	1.00	0.00	0.00	1.97	1.97	0.00	1.95	3.81	0.02
3	2.00	0.00	4.00	0.00	0.00	1.95	3.89	0.00	1.94	3.75	0.01
124	3.00	0.00	9.00	0.00	0.00	1.92	5.76	0.00	1.92	3.69	-0.00
125	4.00	0.00	16.00	0.00	0.00	1.89	7.56	0.00	1.91	3.64	-0.02
126	5.00	0.00	25.00	0.00	0.00	1.86	9.32	0.00	1.89	3.58	-0.03
$\Sigma$	<b>819.0</b>	<b>504.0</b>	<b>7371.</b>	<b>2856.0</b>	<b>3276.0</b>	<b>237.2</b>	<b>1511.3</b>	<b>951.4</b>	<b>237.20</b>	<b>447.00</b>	

## 5. Вимоги до звіту

У звіті слід відобразити мету роботи, коротке теоретичне обґрунтування і привести докладне рішення запропонованих завдань, спираючись на приклад.

## **6. Контрольні питання**

1. Що є предметом тренд-аналізу?
2. Поясніть загальну процедуру виділення тренду.
3. Яким методом визначаються коефіцієнти рівняння тренд-аналізу?
4. Як здійснюється поділ поля досліджуваних геологічних параметрів на локальну і регіональну складові?
5. Дати визначення терміну «аномалія геологічного параметра».
6. Що являє собою тренд першого порядку?
7. Як оцінюється ступень наближення тренду до вихідних даних?

## **Лабораторна робота №4**

### **Визначення характеристик випадкових функцій по одній реалізації**

#### **1. Мета роботи**

Закріплення розуміння особливостей випадкових функцій. Засвоєння методики та прийомів визначення характеристик випадкової функції по одній реалізації.

#### **2. Теоретичні основи**

**Поняття стаціонарної випадкової функції.** Зі стаціонарними випадковими функціями дуже часто доводиться стикатися на практиці, особливо при вивченні фізичних і технічних явищ. Стаціонарні випадкові функції характеризуються, насамперед, однорідністю протікання в часі. Ця ознака і є основною при визначенні, стаціонарна досліджувана функція чи ні. Як правило, процеси, що протікають в системах, що працюють тривалий час, описуються стаціонарними випадковими функціями.

Припустимо, що в результаті проведення дослідів отримано кілька реалізацій випадкової функції та знайдено оцінки математичного сподівання, дисперсії та кореляційної функції. Використовуючи теорію статистичної перевірки гіпотез, можна з'ясувати, випадкові чи ні зміни середніх арифметичних і емпіричних дисперсій ординат і зміни оцінок кореляційних моментів при фіксованій різниці моментів часу, для яких взяті ординати.

Якщо зміни носять випадковий характер, то математичне сподівання і дисперсію випадкової функції можна вважати постійними величинами, а кореляційну функцію залежною тільки від проміжку часу між ординатами, тобто випадкова функція стаціонарна.

Однак, при вирішенні практичних задач, як правило, дослідник обмежений

невеликим числом реалізацій випадкової функції. Це може призвести до того, що теорія статистичної перевірки гіпотез, що припускає досить велику кількість спостережень, не підтвердить гіпотези про стаціонарність функції, хоча природа явища дозволяє судити про те, що воно має описуватися стаціонарною випадковою функцією. Тому вирішити, стаціонарна чи ні досліджувана функція, можна тільки вивчивши внутрішню структуру досліджуваного явища.

Метод визначення характеристик стаціонарної випадкової функції досить складний і включає два етапи: оцінки характеристик ординат випадкової функції з реалізації та осереднення цих оцінок. Природно виникає питання: чи не можна для стаціонарної випадкової функції цей двоступеневий процес замінити більш простим, припустивши, що єдина реалізація достатньої тривалості є достатньо надійним матеріалом для отримання характеристик випадкової функції? При більш детальному розгляді цього питання виявляється, що не для всіх стаціонарних функцій одна реалізація еквівалентна множині реалізацій. Розглянемо, наприклад, дві стаціонарні випадкові функції (див. Рис. 4.1 і 4.2).

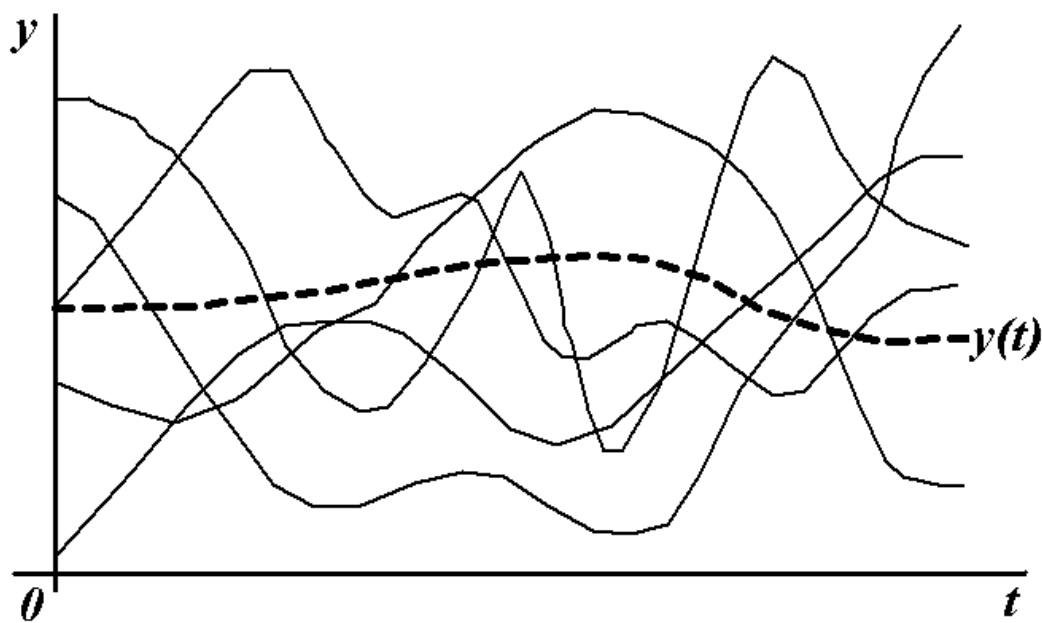


Рис. 4.1 Графік стаціонарної випадкової функції, що має ергодичну властивість

Кожна реалізація випадкової функції, зображеної на рис. 4.1, володіє одними і тими ж характерними ознаками. Очевидно, що одна з таких довільно вибраних реалізацій при досить великому часі випробування зможе дати нам досить гарне уявлення про властивості випадкової функції. Зокрема, усереднивши значення цієї реалізації за часом, можна отримати наближене значення математичного очікування випадкової функції. Про випадкову функцію, кожна досить тривала реалізація якої є як би «повноважним представником» всієї сукупності можливих реалізацій, кажуть, що вона володіє ергодичною властивістю.

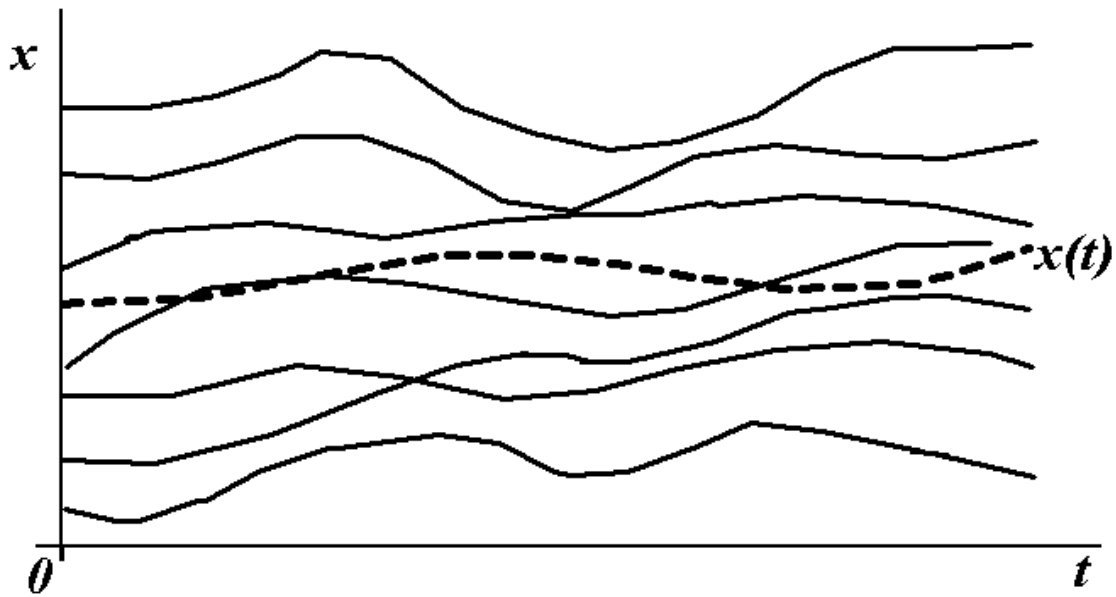


Рис. 4.2 Графік стаціонарної випадкової функції, що не має ергодичної властивості

Розглянемо випадкову функцію, зображену на рис. 4.2. Виберемо довільно одну реалізацію, продовжимо її подумки на досить тривалий проміжок часу і усереднимо її значення. Очевидно, що середня для кожної реалізації може значно відрізнятись від математичного очікування випадкової функції. Про таку випадкову функцію кажуть, що вона не має властивості ергодичності.

Отже, якщо стаціонарна випадкова функція  $X(t)$  володіє ергодичною властивістю, то:

1) її математичне очікування  $a$  наближено дорівнює середній за часом ординаті однієї довільно взятої реалізації  $x(t)$  достатньої тривалості, тобто

$$a \approx \frac{1}{T} \int_0^T x(t) dt; \quad (4.1)$$

2) значення кореляційної функції  $K(\tau)$  при будь-якому  $\tau$  наближено дорівнює середньому за часом добутку відхилень ординат реалізації  $x(t)$  в точках, віддалених один від одного на величину  $\tau$ , від математичного очікування стаціонарної випадкової функції  $a$ , тобто

$$K(\tau) \approx \frac{1}{T - \tau} \int_0^{T-\tau} [x(t) - a][x(t + \tau) - a] dt. \quad (4.2)$$

Тут взято середнє значення по інтервалу  $(0, T - \tau)$ , так як функції  $x(t)$  і  $x(t + \tau)$  відомі сумісно тільки в цьому інтервалі. Формула (4.2) дає оцінку кореляційної функції при  $\tau > 0$ . При  $\tau < 0$  оцінка кореляційної функції



визначається з умови її парності. Формулою (4.2) рекомендується користуватися при  $\tau < T/5$ . При великих значеннях  $\tau$  похибка оцінки кореляційної функції за формулою (4.2) збільшується.

Якщо  $\tau = 0$ , то  $K(0) = \sigma^2$  і дисперсія стаціонарної випадкової функції, що володіє властивістю ергодичності, може бути наближено визначена за формулою

$$\sigma^2 \approx \int_0^T [x(t) - a]^2 dt. \quad (4.3)$$

**Визначення характеристик ергодичної стаціонарної випадкової функції по одній реалізації з дослідження.** Обчислюючи характеристики ергодичної стаціонарної випадкової функції по одній реалізації, використовуючи формули, наведені вище, ми припускали, що аргумент випадкової функції  $t$  може приймати будь-які значення на інтервалі  $(0, T)$  і що заздалегідь відомий аналітичний вигляд реалізації  $x(t)$ . При вирішенні практичних задач в результаті експерименту, як правило, виявляються відомими значення реалізації  $x(t)$  для кінцевої множини значень аргументу і при обчисленні характеристик ергодичної стаціонарної випадкової функції інтеграли (4.1) і (4.2) замінюють кінцевими сумами. Покажемо, як це робиться. Розіб'ємо інтервал  $(0, T)$  на  $n$  рівних частин, довжина яких  $\Delta t$ , і позначимо середини отриманих ділянок через  $t_1, t_2, \dots, t_n$  (рис. 4.3).

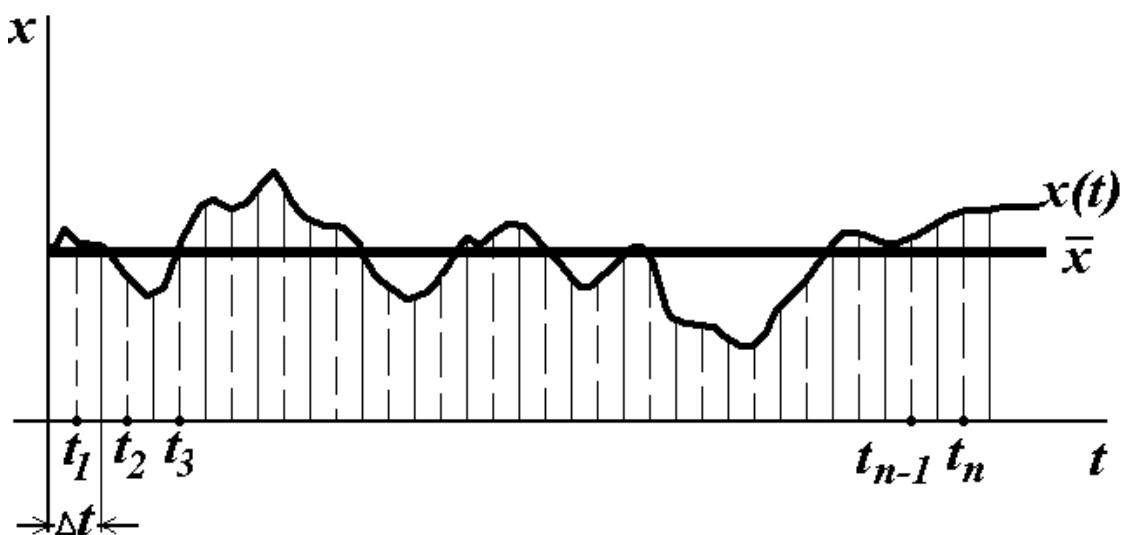


Рис. 4.3. Ілюстрація заміни інтегрування сумою

Будемо розглядати інтеграл (4.1) як суму криволінійних трапецій з висотою, що дорівнює  $\Delta t = T/n$ , і середньою лінією, що рівна  $x(t_i)$ . тоді маємо

$$a \approx \frac{1}{T} \cdot \frac{T}{n} \cdot \sum_{i=1}^n x(t_i),$$

або

$$a \approx \frac{1}{n} \cdot \sum_{i=1}^n x(t_i) = \bar{x}. \quad (4.4)$$

Аналогічно можна обчислити кореляційну функцію при  $\tau = 0, \Delta t, 2\Delta t, \dots$ . Прийmemo, наприклад,  $\tau = m\Delta t = (mT)/n$ . Обчислимо інтеграл (4.2), розділивши інтервал інтегрування, довжина якого дорівнює

$$T - \tau = T - \frac{mT}{n} = \frac{n-m}{n} T,$$

на  $(n-m)$  рівних ділянок довжиною  $\Delta t$ . При цьому заміниmo  $a$  його оцінкою  $\bar{x}$ . Маємо

$$K(m\Delta t) = K\left(\frac{mT}{n}\right) \approx \frac{1}{\frac{n-m}{n} \cdot T} \cdot \frac{T}{n} \sum_{i=1}^{n-m} [x(t_i) - \bar{x}][x(t_{i+m}) - \bar{x}]$$

або

$$K\left(\frac{mT}{n}\right) \approx \frac{1}{n-m} \sum_{i=1}^{n-m} [x(t_i) - \bar{x}][x(t_{i+m}) - \bar{x}] = \tilde{K}\left(\frac{mT}{n}\right). \quad (4.5)$$

При  $m=0$  отримаємо

$$\sigma^2 \approx \frac{1}{n} \sum_{i=1}^{n-m} [x(t_i) - \bar{x}]^2 = s^2. \quad (4.6)$$

Для того щоб математичне сподівання  $a$  і кореляційна функція  $K(\tau)$  були визначені з достатньою точністю, потрібно щоб число точок  $n$  було досить велике. Кількість точок визначається в залежності від властивостей випадкової функції: для функції, що змінюється порівняно плавно, можна обмежитися меншим числом точок, ніж для функції, що має різкі і часті коливання.

**Методика визначення характеристик випадкової функції по одній реалізації.** Обчислення оцінки математичного очікування ергодичної стаціонарної випадкової функції за формулою (4.4) можна розглядати як згладжування її реалізації, отриманої в результаті досліду (рис. 4.3).

Математичне сподівання нестационарної випадкової функції можна також визначити згладжуванням однієї її реалізації (рис. 4.4), якщо отримана в результаті досліду реалізація випадкової функції характеризується достатньою тривалістю і добре представляє всю сукупність можливих реалізацій. Для згладжування користуються методом ковзного середнього. Цей метод полягає в тому, що за згладжене значення функції в будь-якій точці  $t$  приймають середнє значення в деякому інтервалі з центром в цій точці. При зміні  $t$  інтервал ковзає уздовж осі  $t$ , чим й пояснюється назва методу.

Розіб'ємо інтервал  $(0, T)$  випадкової функції на рівні частини точками  $0 = t_1, t_2, \dots, t_k, \dots, t_n = T$ . Для визначення згладженого значення функції в точці  $t_k$  виділимо з множини точок  $t_1, t_2, \dots, t_k, \dots, t_n$  таку підмножину  $(2p + 1)$  точок  $t_{k-p}, \dots, t_{k-1}, t_k, t_{k+1}, \dots, t_{k+p}$ , в якій центральною точкою є точка  $t_k$ . Згладжене значення або, інакше кажучи, значення ковзного середнього  $(2p + 1)$ -го порядку в точці  $t_k$  обчислюється за формулою

$$\bar{x}(t_k) = \frac{x(t_{k-p}) + \dots + x(t_{k-1}) + x(t_k) + x(t_{k+1}) + \dots + x(t_{k+p})}{2p + 1} \dots \dots \dots (4.7)$$

Використовуючи формулу (4.7), можна обчислити згладжені значення функції в точках  $t_{p+1}, t_{p+2}, \dots, t_{n-p}$ .

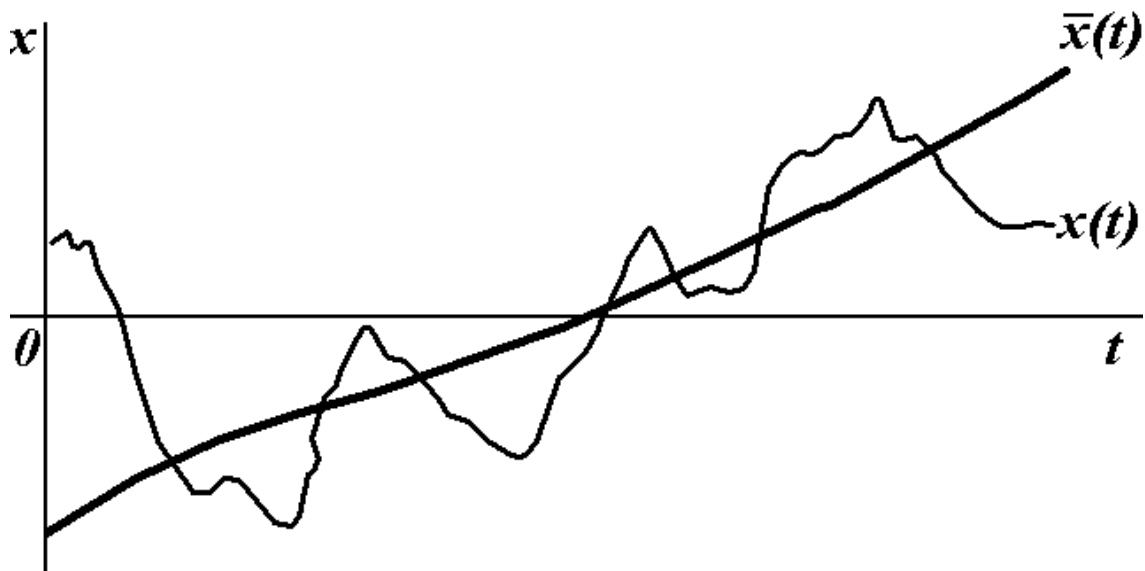


Рис. 4.4. Випадкова функція  $x(t)$  та графік її середнього  $\bar{x}(t)$ , отриманого методом ковзного середнього

Як вибрати інтервал згладжування або, інакше кажучи, яким повинен бути порядок ковзної середньої? Очевидно, чим більше інтервал, тим сильніше згладжування. Але, з іншого боку, при дуже великому інтервалі згладиться і

саме математичне сподівання випадкової функції. Тому інтервал рекомендується вибирати таким, щоб при будь-якому його розташуванні в межах інтервалу  $(0, T)$  реалізація випадкової функції мала б на ньому досить велику кількість коливань і, в той же час, щоб математичне сподівання випадкової функції на інтервалі можна було б вважати приблизно лінійним.

### **3. Порядок виконання роботи**

1. Побудувати графік випадкової функції. Зробити висновок щодо її стаціонарності.
2. Оцінити необхідну кількість значень в ковзному «вікні».
3. Отримати згладжені значення за формулою (4.7). Проаналізувати поведінку згладженою функції.
4. Побудувати графік відхилень вихідної функції від згладжених значень.
5. Обчислити за формулами (4.5) і (4.6) кореляційну функцію і дисперсію.
6. Побудувати графік кореляційної функції. Зробити висновок про її характер.

### **4. Вимоги до звіту**

У звіті слід відобразити мету роботи, коротке теоретичне обґрунтування і привести докладне рішення запропонованих завдань.

### **5. Контрольні питання**

1. Що означає стаціонарність функції?
2. Що означає ергодичність функції?
3. Чим характеризується кореляційна функція стаціонарного ергодичного процесу?
4. В чому полягає метод «ковзного середнього»?
5. З якою метою виконують згладжування реалізації випадкової функції?
6. Як оцінюють необхідну кількість значень в ковзному «вікні»?

## **Лабораторна робота №5**

**Розподіл дискретних сукупностей. Оцінка вірогідності відкриття родовища і стратиграфічної приуроченості нафтогазопроявів**

### **1. Мета роботи**

Освоєння методики аналізу дискретних сукупностей геологічних даних, встановлення закону їх розподілу. Вірогіднісно-статистичний аналіз дискретних сукупностей геологічних даних.

## 2. Теоретичні основи

Якщо виконується декілька випробувань і вірогідність події  $A$  в кожному випробуванні не залежить від результатів інших випробувань, то такі випробування називають незалежними щодо події  $A$ .

В різних незалежних випробуваннях подія  $A$  може мати або різні вірогідності, або одну і ту ж вірогідність. Будемо далі розглядати лише такі незалежні випробування, в яких подія  $A$  має одну і ту ж вірогідність.

Нижче скористаємося поняттям складної події, розуміючи під ним поєднання декількох окремих подій, які називають простими.

Нехай проводиться  $n$  незалежних випробувань, у кожному з яких подія  $A$  може з'явитися або не з'явитися. Домовимося вважати, що вірогідність події  $A$  в кожному випробуванні одна і та ж та дорівнює  $p$ . Отже, вірогідність ненастання події  $A$  в кожному випробуванні також постійна і дорівнює  $q = 1 - p$ .

Поставимо перед собою задачу обчислити вірогідність того, що при  $n$  випробувань подія  $A$  здійсниться (настане) рівно  $k$  разів і, отже, не здійсниться  $n - k$  разів. Важливо підкреслити, що не потрібно, щоб подія  $A$  повторилася рівно  $k$  раз в певній послідовності. Наприклад, якщо мова йде про появу події  $A$  три рази в чотирьох випробуваннях, то можливі наступні складні події:  $AAAA$ ,  $A\bar{A}AA$ ,  $\bar{A}AAA$ ,  $\bar{A}A\bar{A}A$ . Запис  $AAAA$  означає, що в першому, другому і третьому випробуванні подія  $A$  настала, а в четвертому випробуванні вона не з'явилася, тобто наступила протилежна подія; відповідний сенс мають і інші записи.

Шукану вірогідність позначимо  $P_n(k)$ . Наприклад, символ  $P_5(3)$  означає вірогідність того, що в п'яти випробуваннях подія з'явиться рівно три рази і, отже, не настане два рази. Поставлену задачу можна вирішити за допомогою формули Бернуллі, виведення якої наводиться нижче.

Вірогідність однієї складної події, що полягає в тому, що в  $n$  випробуваннях подія  $A$  настане  $k$  раз і не настане  $n - k$  разів, по теоремі множення вірогідностей незалежних подій дорівнює  $p^k q^{n-k}$ . Таких складних подій може бути стільки, скільки можна скласти сполучень з  $n$  елементів по  $k$  елементів, тобто  $C_n^k$ .

Так як ці складні події несумісні, то за теоремою додавання вірогідностей несумісних подій шукана вірогідність дорівнює сумі вірогідностей всіх неможливих складних подій. Оскільки ж вірогідності всіх цих складних подій однакові, то шукана вірогідність (поява  $k$  раз події  $A$  в  $n$  випробуваннях) дорівнює вірогідності однієї складної події, помноженої на їх число:

$$P_n(k) = C_n^k p^k q^{n-k} \quad (5.1)$$

або

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad (5.2)$$

Отриману формулу називають формулою Бернуллі.

У геології за цією формулою розраховують вірогідність прояву одиничного значення досліджуваної ознаки у випадку, коли дискретні сукупності можуть бути описані біноміальним розподілом. У цій формулі  $p$  – параметр розподілу,  $n$  – обсяг сукупності,  $k$  – поточне значення ознаки.

Стандартне відхилення сукупності

$$\sigma_B = \sqrt{np(1-p)} \quad (5.3)$$

Біноміальний розподіл характеризується єдиним параметром  $p$ , а графічне зображення цих вірогідностей являє собою не безперервну криву, а лише систему ординат, пропорційних значенням вірогідностей дискретних значень ознаки.

При вирішенні геологічних задач використовується ще один дискретний розподіл – розподіл Пуасона. Особливістю цього розподілу є можливість кількісної оцінки вірогідності події при малих значеннях  $p$  (при  $p \rightarrow 0$  біноміальний розподіл втрачає сенс).

Вірогідність таких подій оцінюється за асимптотичною формулою Пуасона.

$$P(k\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (5.4)$$

де  $P(k\lambda)$  – вірогідність;  $\lambda$  – одиничний параметр розподілу, що розраховується за формулою

$$\lambda = \bar{h} = \sigma^2 = Np \quad (5.5)$$

Остання обставина є характерною рисою цього розподілу:  $\bar{h} = \sigma^2$  сукупність може бути описана формулою Пуасона.

В практиці геологічних досліджень у якості випадкової події часто виступають факти відкриття родовищ вуглеводнів (ВВ). Відповідно, показником ефективності розвідки виступає ймовірність відкриття родовища при бурінні свердловини, або ймовірність прояву ВВ. Такі ймовірності можна отримати на основі аналізу результатів буріння свердловин. Однак при цьому важливо визначити характер розподілу проявів ВВ, що виступають як

випадкові події (у поточній роботі передбачено аналіз вихідних даних на відповідність законам Бернуллі або Пуассона).

### 3. Завдання та порядок виконання роботи

**Завдання 1.** Використовуючи біноміальний розподіл, за даними буріння і нафтогазопромислової геології слід визначити оптимальну кількість свердловин, необхідних для відкриття родовища ВВ.

**Завдання 2.** Використовуючи розподіл Пуассона, оцінити вірогідність нафтогазопроявів з досліджуваного комплексу гірських порід. Виконати порівняння практичних і теоретичних частот нафтогазопроявів.

*Приклад рішення завдання 1.* На території декількох суміжних площ Дніпровсько-Донецької западини при випробуванні карбонатно-глинистих відкладень московського ярусу в 38 свердловинах було отримано 12 промислових припливів нафти.

Оцінюючи орієнтовно вірогідність розкриття нафтовмістовної частини досліджуваних відкладень через  $p \approx 0,32$ , а непродуктивної – через,  $(1 - p) \approx 0,68$  за формулою (5.2) проведемо розрахунок вірогідності отримання промислового припливу нафти в стратиграфічному інтервалі, що представляє потенційний інтерес (таблиця 5.1) при деяких  $k$  значеннях.

На графіку розподілу (рис. 5.1) видно, що максимальна вірогідність розкриття продуктивної частини досліджуваного стратиграфічного інтервалу відповідає моменту введення в розвідку 10–12 свердловин. Причому «пік» максимальної вірогідності досить гострий – значення стандарту  $\sigma_B \approx 2,87$ .

Таблиця 5.1.

Вірогідність розкриття нафтовмістовного інтервалу при випробуванні карбонатно-глинистих відкладень Московського ярусу ДДЗ

Кількість свердловин, $k$	Вірогідність, $P_n(k)$	Кількість свердловин, $k$	Вірогідність, $P_n(k)$
4	0,001828887	14	0,1050817
6	0,01457048	16	0,05148386
8	0,05498119	18	0,01655798
10	0,1132157	20	0,003527143
12	0,1381242	22	0,0004,976435

*Приклад рішення задачі 2.* Розглянемо особливості розподілу Пуассона на прикладі розподілу нафтогазопроявів, отриманих під час випробування відкладень Башкирського ярусу Дніпровсько-Донецької западини. Аналіз гістограми (рис. 5.2) дозволив виділити п'ять основних продуктивних пластів (III, IV, V, VI, VII), з якими пов'язана більшість нафтогазопроявів, і в тому числі

практично промислові притоки нафти і газу. Тому, якщо врахувати, що інші пласти і пропластки практично виконують роль покришок-екранів, а нафтогазопрояви в них випадкові, нехарактерні в регіональному плані прояви, можна скласти ряд дискретних сукупностей, що характеризують нафтогазопрояви тільки в основних продуктивних пластах.

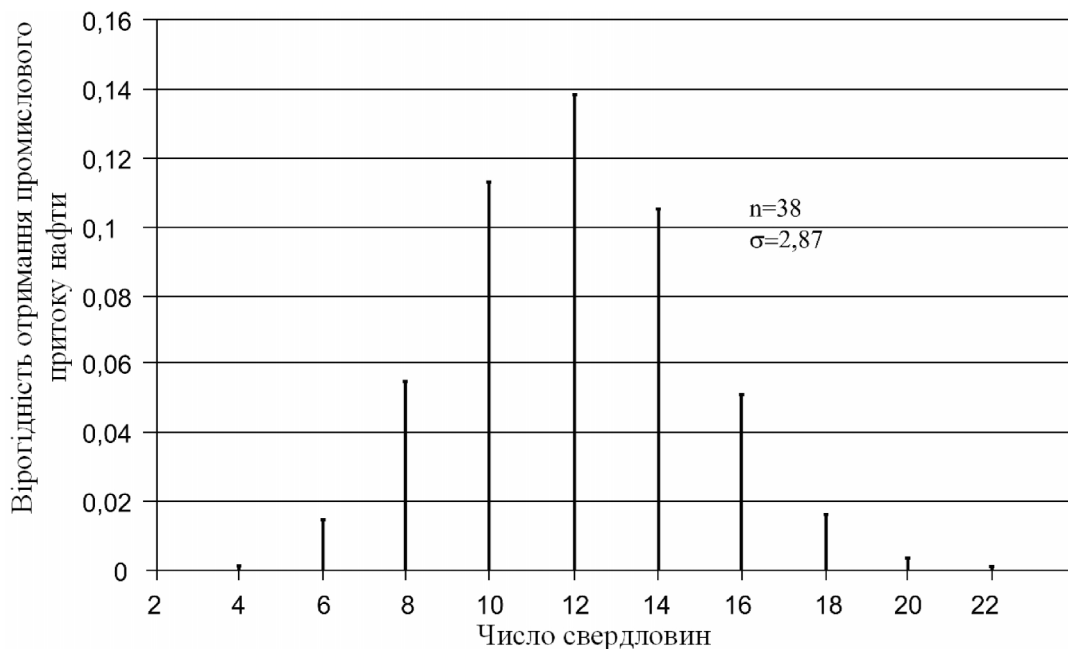


Рис. 5.1. Біноміальний розподіл на прикладі вивчення вірогідності розкриття нафтовмістовної частини карбонатно-глинистих відкладень Московського ярусу ДДЗ

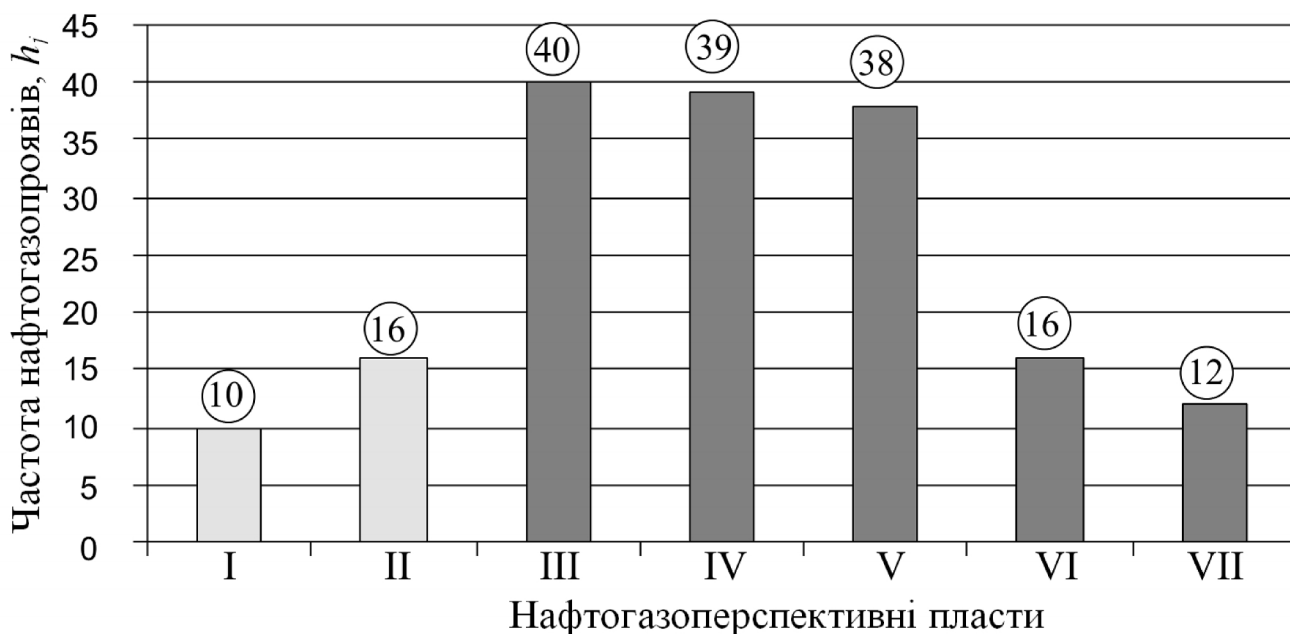


Рис. 5.2. Гістограма нафтогазопроявів при випробуванні відкладів Башкирського ярусу ДДЗ



Вірогідність нафтогазопроявів в кожному з комплексів можна оцінити за формулою Пуассона. Для цього необхідно впевнитися в тому, що розглянута сукупність відповідає умові (5.5), тобто в рівності середнього і дисперсії. Слід зазначити, що практично така рівність при вивченні сукупностей геолого-геофізичних даних нездійсненна, але можна дотримуватися умови  $\bar{h} \approx \sigma^2$ . Результати розрахунку зведено у таблицю (5.2).

Таблиця 5.2.

Розрахунок параметрів сукупності нафтогазопроявів з відкладень Башкирського ярусу ДДЗ

Продуктивний пласт	Кодовий номер комплексу, $h_j$	Частота нафтогазопроявів, $f_j$	$h_j f_j$	$h_j - \bar{h}$	$(h_j - \bar{h})^2$	$f_j (h_j - \bar{h})^2$
1	2	3	4	5	6	7
III	0	40	0	-0,7158273	0,5124088	20,49635
IV	1	39	39	0,2841727	0,0807541	3,14941
V	2	32	64	1,284173	1,649099	52,77118
VI	3	16	48	2,284173	5,217445	83,47912
VII	4	12	48	3,284173	10,78579	129,4295
		$\sum f_j = N$	$\sum h_j f_j$			$\sum f_j (h_j - \bar{h})^2$
		139	199			289,3255

Звідси значення середнього  $\bar{h} = \frac{199}{139} = 0,7158273$  і дисперсії

$\sigma^2 = \frac{289,3255}{139} = 1,040739$ . Будемо вважати їх допустимо близькими, щоб дана сукупність могла бути описана розподілом Пуассона. Обчислимо значення вірогідностей по (5.4):

при  $k=0$

$$P(0; 0,89) = 0,410655753;$$

при  $k=1$

$$P(1; 0,89) = 0,36548362;$$

при  $k=2$

$$P(2; 0,89) = 0,162640211;$$

при  $k=3$

$$P(3; 0,89) = 0,048249929;$$

при  $k=4$

$$P(4; 0,89) = 0,010735609.$$

Обчислені за формулою Пуассона значення вірогідностей можуть бути використані для розрахунку теоретичних значень частот  $F_k$  за формулою:

$$F_k = P(k = j, \lambda) \sum_{j=0}^n f_j \quad (5.6)$$

Наприклад, теоретична частота нафтогазопроявів в III пласті Башкирського ярусу ( $k = 0$ ) дорівнює  $F_0 = 0,410655753 \times 139 \approx 57$  при фактичній частоті, що складає 40.

#### **4. Вимоги до звіту**

У звіті слід відобразити мету роботи, коротке теоретичне обґрунтування і привести докладне рішення запропонованих завдань спираючись на приклади.

#### **5. Контрольні питання**

1. Що визначає формула Бернуллі?
2. Що є єдиним параметром біномного розподілу?
3. Яка основна умова застосування біноміального розподілу?
4. Яка основна умова застосування розподілу Пуассона?
5. Як визначаються теоретичні та фактичні частоти?
6. В чому полягає практична цінність зображення розподілу величин у вигляді гістограм?