

Розділ 6. Регресійний аналіз

6.1. Комп'ютер приймає участь в експерименті

Експеримент завжди був основою наукових досліджень в області хімії та хімічної технології. Йде час, ускладнюється експеримент, зростає об'єм інформації, що одержують в результаті їх проведення. Досліднику важче стає аналізувати дані в ході експерименту та керувати складними експериментальними установками. Ці обставини зумовили необхідність створення **автоматизованих систем наукових досліджень (АСНД)**. В основі роботи кожної такої системи – використання ЕОМ, але принципи використання комп'ютерів можуть бути різними.

У найпростіших випадках під час проведення експерименту здійснюється тільки збір інформації шляхом вимірювання та запису різних фізичних величин. Обробка та аналіз даних здійснюється після завершення експерименту. Такі системи використовуються тільки для нескладних та повільних процесів. Окрім того, управління експериментом дослідник проводить сам.

Зустрічаються більш складні умови проведення експерименту. По-перше, як правило, хімічні процеси протікають дуже швидко. По-друге, обробку експериментальних даних потрібно проводити під час проведення дослідів. По-третє, є необхідність автоматизованого управління перебігом експерименту. Для вирішення цих складних задач потрібна автоматизована система, що управляється комп'ютером.

Якщо розрахунки проводяться попередньо або здійснюються після проведення експерименту з метою обробки одержаних даних, то можливості сучасних ЕОМ дозволяють сподіватися, що з цими розрахунками можна справитися, якщо використовувати традиційні методи організації та управління процесами. Але при проведенні розрахунків за допомогою ЕОМ під час експерименту з'являються специфічні проблеми в управлінні розрахунками та створенні відповідних програм.

Найбільшої ефективності експерименту можна досягти, якщо обробляти дані в темпі їх одержання при проведенні дослідів. Високі можливості ЕОМ дозволяють сподіватися, що всі особливості явища, що вивчається, будуть зафіксовані, а при відхиленні процесу від наперед запланованого режиму, існує можливість вчасно скоректувати його. Сигнали керування виробляє одна з програм, що записана в пам'ять ЕОМ. Таким чином, ЕОМ стає залученою до схеми управління експериментом (див. рис.6.1.1).

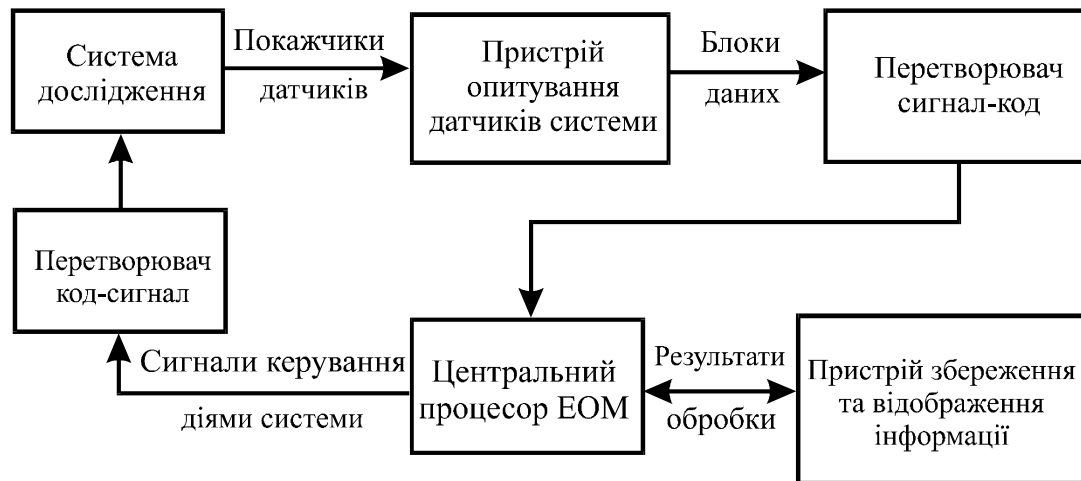


Рис.6.1.1.Схема використання ЕОМ при проведенні експериментальних досліджень.

Від датчиків експериментальної установки сигнали різної фізичної природи в певний момент часу, що задається пристроєм опитування датчиків, надходить на вхід перетворювача сигнал–код. Такий перетворювач називають аналогово-цифровим. На його вихідних регістрах формуються числові значення відповідних фізичних величин у вигляді цифрових записів, що можуть бути оброблені на ЕОМ. За допомогою каналів зв'язку обчислювальної машини ці дані надаються для обробки в оперативну пам'ять. Виміри, що відносяться до одного моменту опитування датчиків, групуються в один кадр. Канали зв'язку ЕОМ мають особливість, в силу якої вигідно якомога більше число кадрів об'єднувати в єдиний блок для передачі в оперативну пам'ять. Це пояснюється тим, що процес пересилки даних здійснюється протягом дуже малого проміжку часу, фактично за час, необхідний для проходження електричними сигналами відстані від вихідних регістрів аналогово-цифрових перетворювачів до регістрів оперативної пам'яті. Значно більший час потрібен для розміщення даних в пам'яті ЕОМ.

Якщо центральний процесор ЕОМ в момент одержання даних виконує команди деякої програми, то відбувається переривання цих розрахунків, проводиться запам'ятовування стану перерваної програми для подальшого продовження її роботи. Процесор надається в розпорядження програми, що контролює надходження нової порції даних. Щоб зменшити витрати машинного часу на зв'язок ЕОМ із об'єктом досліджень, сеанси зв'язку вигідно проводити не дуже часто, а для цього потрібно укрупнювати блоки даних, об'єднувати коди, що відносяться до різних моментів часу, в єдину упорядковану групу.

Таким чином, навіть простий прийом даних експерименту потребує узгодження роботи ЕОМ із процесом одержання даних від експериментальної установки. Одержані оперативною пам'яттю ЕОМ дані обробляються програмами, що розраховують різні характеристики досліджуваного під час експерименту явища. Одні результати цих розрахунків повинні оперативно

виводитися на пристрої відображення інформації – алфавітно-цифрові та графічні дисплеї, друкуючі пристрої, другі – запам’ятовуються з метою подальшого більш ретельного аналізу, треті – використовуються в якості вхідних значень для інших програм, що застосовують для розрахунків нових характеристик. Тому дії обчислювальної системи повинні бути синхронізовані з зовнішніми процесами, тобто ЕОМ повинна працювати в режимі реального часу. Навіть незначна затримка у видачі деяких програм управління буде сприйматися як відмова системи в роботі. Тому надійність – головна вимога при створенні систем реального часу. Вона забезпечується шляхом підвищення надійності апаратних засобів, а також відповідних алгоритмів та програм.

6.2. Оптимізація перебігу хімічної реакції (модельна задача)

Розглянемо схему хімічної реакції, що складається з декількох стадій. Механізм перебігу реакції включає ряд послідовних та паралельних стадій:

де А – вихідна речовина, В – проміжковий продукт, С – кінцевий продукт, D – небажані продукти.

Залежність константа швидкості (k_i) окремих i -тих стадій реакції від температури описується рівнянням Арреніуса:

$$k_i = k_{0i} e^{\frac{-E_i}{RT}} \quad (6.2.1),$$

де k_i – константа швидкості i -ої стадії; k_{0i} – передекспоненційний множник; E_i – енергія активації i -ої стадії; R – універсальна газова стала; T – температура.

Будемо рахувати, що хімічна реакція проходить у трубчатому реакторі з ідеальним витісненням реакційної маси.

Рис.6.2.1. Схема перебігу реакції в трубчатому реакторі

Слід зауважити, що кінцевий продукт С спочатку утворюється з речовини В, а потім переходить в речовину D. Це вказує на те, що якщо час перебування реакційної маси в реакторі дуже малий, то і концентрація речовини С на виході із реактора (в потоці) буде незначна. Із збільшенням часу перебування реакційної маси в реакторі концентрація продукту С буде зростати. Проте, починаючи з певного моменту часу, концентрація речовини С почне зменшуватись внаслідок перетворення її у небажаний продукт D. Отже існує певне значення часу перебування реакційної суміші в трубчатому реакторі, при якому концентрація речовини С на виході потоку буде максимальною. Це значення часу перебування суміші в апараті називається оптимальним.

Температура також суттєво впливає на працездатність реактора за кінцевим продуктом. Тому задача оптимізації полягає в тому, щоб знайти розподіл температури вздовж трубчатого реактору, який максимізує концентрацію кінцевої речовини С у вихідному потоці.

Для оптимізації технологічного процесу проходження хімічної реакції побудуємо математичну модель трубчатого реактору, використовуючи закон діючих мас та принцип незалежності перебігу окремих стадій хімічних реакцій. Дана модель має вигляд системи звичайних диференціальних рівнянь з початковими значеннями $x_1(0)=1$, $x_2(0), x_3(0)=0$:

$$\frac{dx_1}{dt} = -(k_1 + k_2 + k_3)x_1 \quad (6.2.2); \quad \frac{dx_2}{dt} = k_1x_1 - k_4x_2 \quad (6.2.3);$$

$$\frac{dx_3}{dt} = k_4x_2 - k_5x_3 \quad (6.2.4),$$

де x_1, x_2, x_3 – відповідно концентрації речовин А, В, С.

Приймаючи, що для невеликого проміжку часу $\frac{dx}{dt} = \frac{\Delta x}{\Delta t}$ та $\Delta x = x(k) - x(n)$, одержуємо:

$$x_1(k) = x_1(n) - (k_1 + k_2 + k_3) \times x_1(n) \times \Delta t \quad (6.2.5)$$

$$x_2(k) = x_2(n) + (k_1 \times x_1(k) - k_4 \times x_2(n)) \times \Delta t \quad (6.2.6)$$

$$x_3(k) = x_3(n) + (k_4 \times x_2(k) - k_5 \times x_3(n)) \times \Delta t \quad (6.2.7)$$

Задача оптимізації полягає в знаходженні розподілу температури (Т) та значення часу перебування суміші в апараті (t), при яких концентрацій речовини С (x_3) буде максимальною. Верхню межу температури ($T_{\text{макс}}$), як правило, обмежують технологічними міркуваннями. Зміна температури (охолодження суміші) вздовж трубчатого хімічного реактору описується рівнянням:

$$T = a1 - a2 \times t \quad (6.2.8),$$

де a_1 – температура суміші на вході в реактор, a_2 – швидкість охолодження, t – час перебування суміші в реакторі. Інтегрування диференціальних рівнянь здійснюється методом Ейлера.

Текст програми на мові Basic .” Оптимізація складного хіміко-технологічного процесу (модель хімічної реакції)”

```

10 REM Модель хімічної реакції
20 DIM c(5), k(5), e(5)
30 INPUT "Введіть початкову температуру та швидкість охолодження:"; a1, a2
40 FOR i = 1 TO 5
50 INPUT "Введіть значення енергії активації (Ei=)"; e(i)
60 NEXT
70 FOR i = 1 TO 5
80 INPUT "Введіть значення k0i="; k0(i)
90 NEXT
100 x1 = 1: x2 = 0: x3 = 0
110 t1 = 0: t2 = 1: d = 0.02
120 t = a1: r = 1.9862
130 k = 0
140 FOR i = 1 TO 5
150 k(i) = k0(i) * EXP(-e(i)) / (r * t)
160 NEXT i
170 x1 = x1 - (k(1) + k(2) + k(3)) * x1 * d
180 x2 = x2 + (k(1) * x1 - k(4) * x2) * d
190 x3 = x3 + (k(4) * x2 - k(5) * x3) * d
200 t1 = t1 + d: t = a1 - a2 * t1
210 PRINT "t(час) ="; t1; TAB(25); "t(температура)="; t; TAB(40); "x3="; x3
220 IF t1 > t2 THEN 270
230 k = k + 1
240 IF k < 11 THEN 140
250 INPUT "„Продовжувати розрахунок – 1, STOP – 2:"; z
260 ON z GOTO 130, 270
270 END

```

Пояснення до програми.

У програмі використовуються змінні та сталі величини: **a_1** – початкове значення температури ($^{\circ}\text{C}$); **a_2** – швидкість охолодження реакційної суміші ($^{\circ}\text{C}/\text{сек}$); **$e(i)$** – енергія активації i -ої стадії реакції; **$k_0(i)$** – передекспоненційний множник i -ої стадії реакції; **$k(i)$** – константа швидкості i -ої стадії реакції **x_1, x_2, x_3** – відповідно концентрації речовин А, В, С; **t** – температура процесу ($^{\circ}\text{C}$); **t_1** – початкове значення часу перебування реакційної суміші в апараті; **t_2** – кінцеве значення часу перебування реакційної суміші в апараті; **d** – прирощення часу; **k** – лічильник.

6.3.Лінійна регресія

У науці **регресією** (рух у напрямку протилежному прогресу – руху вперед) називається статистичний зв'язок між випадковими величинами.

Лінійна функціональна залежність між величинами у хімії зустрічається дуже часто. Так, при певних умовах лінійними будуть залежність логарифму тиску пари речовини від оберненої температури, логарифму константи рівноваги реакції від оберненої температури та інші. При цьому параметри лінійних функціональних залежностей визначають важливі фізико-хімічні величини, такі як теплота випаровування, ентальпія, ентропія реакції та інші.

При обробці експериментальних даних хіміки часто будують різні графіки. У багатьох випадках спостерігається лінійна залежність між вимірюваними величинами **X** та **Y**, тому експериментальні точки групуються навколо деякої прямої лінії (рис.6.3.1).

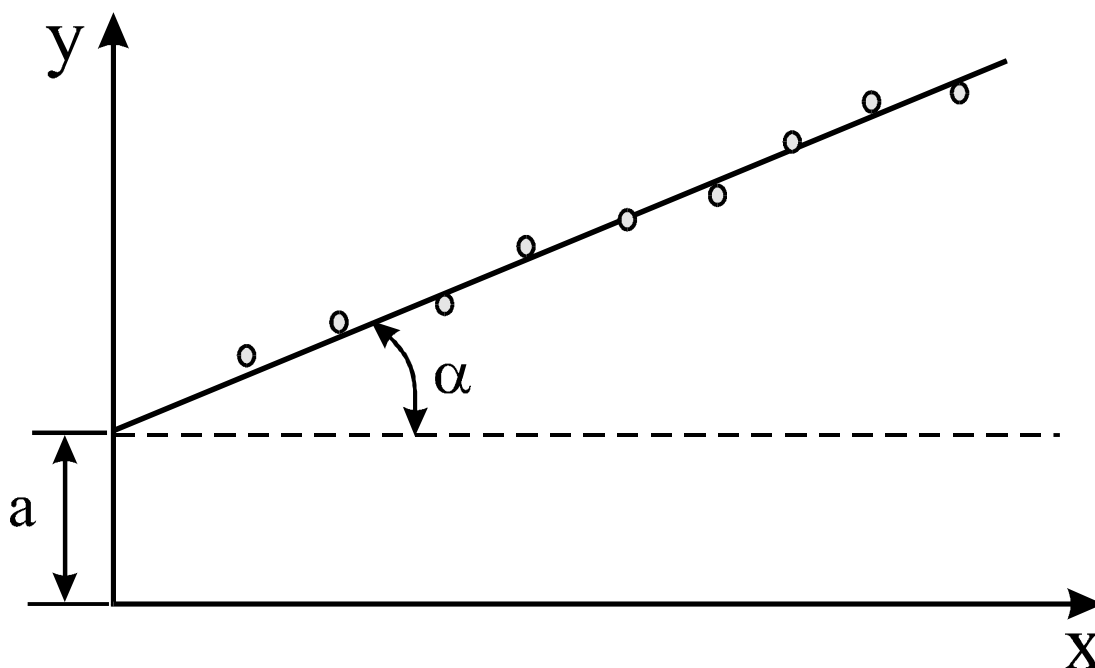


Рис.6.3.1.Лінійна залежність між експериментальними даними

Рівняння прямої лінії має вигляд:

$$Y = a + b \times X \quad (6.3.1),$$

де **a** – довжина відрізка від початку координат до точки перетину прямої з віссю **OY**; **b** – тангенс кута нахилу прямої до осі **OX**.

Пряму лінію намагаються провести таким чином, щоб сума квадратів відхилень експериментальних значень **Y_e** від розрахункових **Y_p** (де **Y_{pi}=a+bX_i**), буде мінімальною для всіх **n** дослідів:

$$S = \sum_{i=1}^n (Y_{ei} - Y_{pi})^2 \quad (6.3.2).$$

Дана процедура відома як метод найменших квадратів (МНК). Кожна ордината Y_i в досліді визначається з певною похибкою, яка характеризується, наприклад, середнім квадратичним відхиленням:

$$Y_i \pm S_y \quad (6.3.3),$$

де S_y – середнє квадратичне відхилення і визначається за формулою:

$$S_y = \sqrt{\frac{S}{n-1}} \quad (6.3.4),$$

Якщо розподіл похибок ординат відповідає нормальному (гауссовому) закону, то при визначенні мінімуму суми квадратів відхилень можливо використовувати відносні вагові коефіцієнти:

$$W_i = 1/S_y \quad (6.3.5),$$

а відхилення використовувати у вигляді $W_i \times (Y_e - Y_p)$.

Даний метод (використовування вагових коефіцієнтів) називається інструментальним зважуванням. Якщо середні квадратичні відхилення невідомі, корисно використовувати статистичне зважування:

$$W_i = 1/Y_i \quad (6.3.6).$$

У багатьох випадках ваговий фактор приймається за одиницю ($W_i=1$).

Математично доведено, що вимозі (6.3.2) задовольняють коефіцієнти рівняння, які розраховуються за формулами:

$$a = \frac{s3 \times s4 - s2 \times s5}{s1 \times s4 - s2 \times s2} \quad b = \frac{s1 \times s5 - s2 \times s3}{s1 \times s4 - s2 \times s2} \quad (6.3.7-8),$$

$$\begin{aligned} s1 &= \sum_{i=1}^n W_i & s2 &= \sum_{i=1}^n W_i X_i & s3 &= \sum_{i=1}^n W_i Y_i \\ s4 &= \sum_{i=1}^n W_i X_i^2 & s5 &= \sum_{i=1}^n W_i X_i Y_i \end{aligned} \quad (6.3.9).$$

У випадку, коли використовується одиничне зважування ($W_i=1$) $S1=n$, де n – загальна кількість дослідів. Рівняння $Y=a+bX$, до якого підставлені розраховані за формулами 6.3.7 та 6.3.8 коефіцієнти називають рівнянням лінійної регресії.

Використовуючи дане рівняння, можливо, не проводячи подальших дослідів, розраховувати для певних значень X відповідні Y , тобто прогнозувати величину Y . Точність прогнозування значення до фактичного залежить від умов, в яких проводиться дослід, та від того, наскільки існуюча залежність між величинами наближується до лінійної.

Текст програми на мові Basic, яка виконує всі необхідні операції, для персональних ЕОМ допоможе точно і швидко розрахувати коефіцієнти регресії.

```

10 CLS : REM Лінійна регресія
20 INPUT "Введіть число експериментальних даних:"; n
30 DIM x(n), y(n)
40 FOR i = 1 TO n
50 PRINT "Введіть значення "; i; "-ої пари X та Y"; : INPUT x(i), y(i)
60 NEXT i
70 s1 = n: s2 = 0: s3 = 0
80 s4 = 0: s5 = 0
90 FOR i = 1 TO n
100 s2 = s2 + x(i): s3 = s3 + y(i)
110 s4 = s4 + x(i) * x(i): s5 = s5 + x(i) * y(i)
120 NEXT
130 d1 = s1 * s4 - s2 * s2
140 d2 = s3 * s4 - s5 * s2
150 d3 = s1 * s5 - s2 * s3
160 a = d2 / d1: b = d3 / d1
170 PRINT "a="; a; TAB(15); " b="; b
180 s = 0
190 FOR i = 1 TO n
200 s = s + (y(i) - (a + b * x(i))) ^ 2
210 NEXT
220 sd = SQR(s / (n - 1))
230 PRINT "Середнє квадратичне відхилення="; sd
240 INPUT "1-здійснити прогноз значень, 2- stop :"; k
250 ON k GOTO 260, 330
260 INPUT "Введіть число значень для прогнозу:"; k1
270 FOR i = 1 TO k1
280 INPUT "Введіть x:"; x
290 y = a + b * x
300 PRINT "y="; y
310 NEXT
320 GOTO 240
330 END

```

Задача. Відомо, що з підвищенням температури розчинність солей у воді зростає. Нагадаємо, що розчинність – це здатність речовин розчинятися у воді та інших розчинниках. Розрахувати коефіцієнти регресії та спрогнозувати розчинність солі при 35 та 55 °С, якщо відомі експериментальні дані про розчинність солі при температурах:

Температура (X), °С:	10	20	30	40	50	60	70	80
Розчинність (Y), г/л:	45	50	62	67	75	86	92	103

Обробка експериментальних результатів за допомогою програми на ЕОМ показала, що рівняння регресії має вигляд

$$Y = 33,71 - 0,85 \times X$$

Середнє квадратичне відхилення $S_y = 1,46$.

Підставляючи в рівняння регресії значення температури ($X=35$ та $X=55^\circ\text{C}$) знаходимо відповідні значення розчинності ($Y=63,46$ та $Y=80,46$ г/л).

6.4. Лінеаризація нелінійних залежностей

У попередній главі (6.2) регресивний аналіз розглядався для випадків, коли лінійна залежність між величинами спостерігається безпосередньо при дослідженні. Але на практиці значно більше випадків, коли зв'язок між двома величинами суттєво нелінійний. Деякі нелінійні залежності можливо звести до лінійних, використовуючи математичні перетворення. Розглянемо найбільш характерні типи нелінійних залежностей.

1. Залежність константи рівноваги реакції від температури для невеликого інтервалу температур має вигляд:

$$\lg K = \frac{S}{4,576} - \frac{H}{4,576} \times \frac{1}{T} \quad (6.4.1),$$

де **S** та **H** – зміна ентропії та ентальпії у процесі реакції. У даному випадку безпосередніми величинами, які спостерігаються під час досліду являються константа рівноваги **K** та температура **T**. З рівняння 6.4.1 бачимо, що зв'язок між ними є нелінійним. Якщо ми проведемо заміну з використанням формул:

$$Y = \lg K \quad a = \frac{S}{4,576} \quad b = -\frac{H}{4,576} \quad X = \frac{1}{T} \quad (6.4.2),$$

одержимо залежність 6.3.1.

Так, використовуючи експериментальну залежність між константою рівноваги **K** та температурою **T**, ми можемо розрахувати значення величин ентальпії та ентропії в ході реакції:

$$H = -4,576 \times b \quad S = 4,576 \times a \quad (6.4.3),$$

де **a** і **b** розраховані за допомогою методу найменших квадратів параметри прямої в координатах **$\lg K = f(1/T)$** .

2. Залежність тиску насиченої пари рідини від температури в невеликому інтервалі визначається рівнянням:

$$\lg p = \text{const} - \frac{H}{4,576} \times \frac{1}{T} \quad (6.4.4).$$

Якщо ввести позначення:

$$Y = \lg p \quad a = \text{const} \quad b = -\frac{H}{4,576} \quad X = \frac{1}{T} \quad (6.4.5),$$

то обробка вимірів тиску пари для визначення теплоти випаровування **H** зводиться до регресивного аналізу залежності $Y=a+bX$. Тому теплота випаровування визначається за формулою 6.3.1.

3. Залежність константи швидкості хімічної реакції від температури реакції описується рівнянням Арреніуса та має експоненціальний характер:

$$k = k_0 \times e^{\frac{-E}{RT}} \quad (6.4.6),$$

де **k** – константа швидкості хімічної реакції, **k₀** – передекспоненційний множник, **E** – енергія активація, **R** – універсальна газова стала, **T** – температура.

Лінеаризація проводиться шляхом логарифмування рівняння 6.4.6, внаслідок якого одержуємо:

$$\ln k = \ln k_0 - \frac{E}{R} \times \frac{1}{T} \quad (6.4.7).$$

Вводимо наступні позначення:

$$Y = \ln k \quad a = \ln k_0 \quad b = -\frac{E}{R} \quad X = \frac{1}{T} \quad (6.4.8).$$

З врахуванням даних позначень одержуємо лінеаризоване рівняння регресії $Y=a+bX$. Розрахувавши коефіцієнти **a** і **b**, знаходимо значення передекспоненційного множника **k₀** та енергії активації:

$$k_0 = e^a \quad E = -b \times R \quad (6.4.9)$$

4. Константа швидкості будь-якої реакції першого ступеня визначається формулою:

$$k = \frac{1}{t} \times \ln \frac{C_0}{C} \quad (6.4.10),$$

де **C₀** – початкова концентрація реагуючої речовини, **C** – її концентрація на момент часу **t**. Перетворивши формулу в рівняння

$$\ln C = \ln C_0 - k \times t \quad (6.4.11)$$

провівши відповідні заміни за допомогою методу МНК, розрахувавши коефіцієнти лінійної регресії для величин константи швидкості хімічної реакції одержуємо наступний вираз:

$$k = -b \quad (6.4.12)$$

Таким чином, при використанні лінеаризації можливо значно розширити область застосування методу лінійного регресивного аналізу. Широке використання нелінійних залежностей та зв'язків, які легко приводяться до лінійних, пояснюється тим, що вони прості та потребують відносно невеликого об'єму розрахунків, а методика їх встановлення розроблена глибше.

6.5. Кореляційний зв'язок між величинами

Встановлюючи зв'язок між величинами, ми наносимо їх на графік. Кожній парі значень величин X та Y на графіку відповідає точка. Як правило, ці точки не знаходяться безпосередньо на прямій, а займають на площині певну область, утворюючи діаграму розсіювання.

Відомо, що зв'язок між двома величинами, які характеризують хіміко-технологічний процес, можна виразити рівнянням регресії. Але воно не дозволяє з абсолютною точністю прогнозувати значення Y , якщо відомо значення X . Прогнозування здійснюється з певною ймовірністю. Зв'язок між двома величинами, який заданий з певною ймовірністю, називають **кореляційним**.

Кореляційний зв'язок між величинами називають **додатним** (рис.6.5.1 а), якщо при збільшенні однієї з них збільшується інша. **Від'ємним** (рис.6.5.1 б) називають кореляційний зв'язок, якщо при збільшенні однієї з величин значення іншої зменшується.

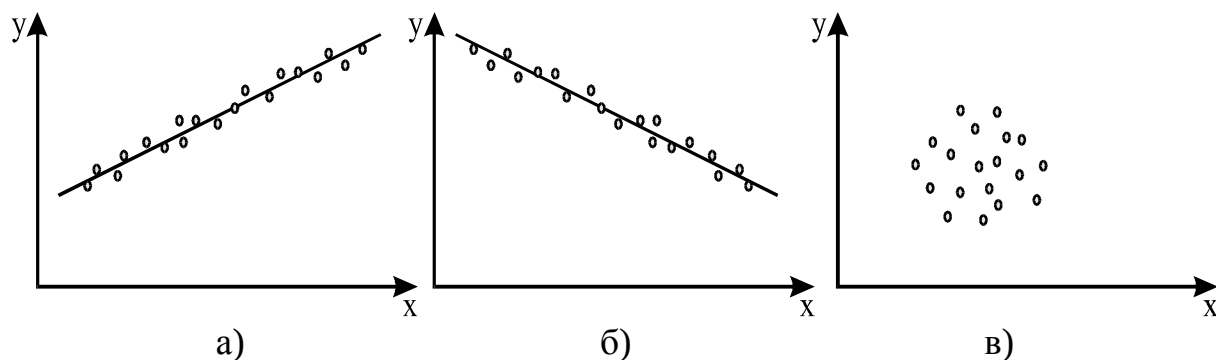


Рис.6.5.1. Діаграма розсіювання: а) додатний зв'язок, б) від'ємний зв'язок, в) відсутність зв'язку

Для кількісної оцінки зв'язку між експериментальними величинами використовують коефіцієнт парної кореляції r_{xy} , який розраховують за рівнянням:

$$r_{xy} = \frac{s1 \times s5 - s2 \times s3}{\sqrt{(s1 \times s4 - s2 \times s2) \times (s1 \times s6 - s3 \times s3)}} \quad (6.5.1),$$

де при одиночному зважуванні

$$\begin{aligned} s1 &= \sum_{i=1}^n n & s2 &= \sum_{i=1}^n X_i & s3 &= \sum_{i=1}^n Y_i & s4 &= \sum_{i=1}^n X_i^2 \\ s5 &= \sum_{i=1}^n X_i Y_i & s6 &= \sum_{i=1}^n Y_i^2 \end{aligned} \quad (6.5.2).$$

Даний коефіцієнт може приймати наступні значення:

- 1) $r_{xy} = 0$; це вказує на відсутність кореляційного зв'язку між величинами X та Y ;
- 2) $r_{xy} = +1$; в даному випадку існує строгий додатній детерміністичний зв'язок;
- 3) $r_{xy} = -1$; в даному випадку існує строгий від'ємний детерміністичний зв'язок;
- 4) $-1 < r_{xy} < +1$ ($r_{xy} \neq 0$); це найбільш поширений випадок: кореляційний зв'язок може бути як додатнім, так і від'ємним та характеризуватися різним ступенем щільності.

Чим ближче за абсолютним значенням коефіцієнт парної кореляції наближується до одиниці, тим лінійний зв'язок між величинами буде сильнішим.

Слід зауважити, що коефіцієнт парної кореляції визначає ступінь випадковості криволінійності зв'язку між величинами X та Y . Наприклад, залежність Y від X може наближуватися до функціональної, але бути суттєво нелінійною. У даному випадку коефіцієнт парної кореляції буде значно меншим за одиницю.

Задача. Дати оцінку кореляційності зв'язку залежності розчинності солі від температури:

Температура (X), $^{\circ}\text{C}$:	10	30	50	70
Розчинність (Y), г/л:	42	62	75	92

Розрахунки за програмою показали, що коефіцієнт парної кореляції r_{xy} дорівнює 0,988. Це відповідає додатному кореляційному зв'язку з великим ступенем щільності.

6.6. Згладжування експериментальних даних

При проведенні експериментальних досліджень спостерігаються часті та різкі зміни значень вимірюваних величин. Це вносить певні труднощі в процес вивчення хімічних явищ. Для уникнення випадкових помилок, що приводять до певного розсіювання значень величин, використовують згладжування експериментальних результатів.

При розрахунках згладженого значення використовують результати вимірювання величин у сусідніх вузлах. Серед великої кількості найбільш широко застосовуються методи простого, три- та п'ятиординатного згладжування. Суттєвим недоліком усіх методів є необхідність експериментального вимірювання величин Y через рівні інтервали ΔX .

1) **Просте згладжування.** При використанні даного методу апріорно приймається, що згладжене значення в першій Y^*_1 та останній Y^*_n точках відповідають експериментальним значенням:

$$Y_1^* = Y_1 \quad (6.6.1) \quad Y_n^* = Y_n \quad (6.6.2).$$

Згладжене значення Y_2^* розраховується за формулою:

$$Y_2^* = 1/5(Y_1 + 3Y_2 + Y_3) \quad (6.6.3).$$

У даному випадку множник 3 називають **ваговим коефіцієнтом**. Він підсилює вплив значення Y_2 на процес згладжування. За аналогічною формулою розраховується значення Y_{n-1}^* :

$$Y_{n-1}^* = 1/5(Y_{n-2} + 3Y_{n-1} + Y_n) \quad (6.6.4).$$

Розрахунок згладжених значень інших величин ($3 \leq i \leq n-2$) здійснюють за формулою:

$$Y_i^* = 1/15(Y_{i-2} + 3Y_{i-1} + 7Y_i + 3Y_{i+1} + Y_{i+2}) \quad (6.6.5).$$

Вагові коефіцієнти 3 та 7 характеризують внесок окремих значень при вивченні згладжуючої функції.

У розглянутому методі враховуються по два сусідніх значення з кожної сторони, яким відповідають вагові коефіцієнти: 1, 3, 7, 3, 1. При необхідності кількість сусідніх значень, які розглядаються, та їх вагові коефіцієнти можливо змінювати.

2) **Триординатне згладжування.** Лінійне згладжування за трьома ординатами проводять за допомогою формул:

$$Y_1^* = 1/6(5Y_1 + 2Y_2 - Y_3) \quad \text{де } i=1 \quad (6.6.6).$$

$$Y_n^* = 1/6(5Y_n + 2Y_{n-1} - Y_{n-2}) \quad \text{де } i=n \quad (6.6.7).$$

$$Y_i^* = 1/3(Y_{i-1} + Y_i + Y_{i+1}) \quad \text{де } 2 \leq i \leq n-1 \quad (6.6.8).$$

3) **П'ятиординатне згладжування.** Лінійне згладжування за п'ятьма ординатами проводять за допомогою формул:

$$Y_1^* = 1/5(3Y_1 + 2Y_2 + Y_3 - Y_4) \quad \text{де } i=1 \quad (6.6.9).$$

$$Y_n^* = 1/5(3Y_n + 2Y_{n-1} + Y_{n-2} - Y_{n-3}) \quad \text{де } i=n \quad (6.6.10).$$

$$Y_2^* = 1/10(4Y_1 + 3Y_2 + 2Y_3 + Y_4) \quad \text{де } i=2 \quad (6.6.11).$$

$$Y_{n-1}^* = 1/10(4Y_n + 3Y_{n-1} + 2Y_{n-2} + Y_{n-3}) \quad \text{де } i=n-1 \quad (6.6.12).$$

$$Y_i^* = 1/5(Y_{i-2} + Y_{i-1} + Y_i + Y_{i+1} + Y_{i+2}) \quad \text{де } 3 \leq i \leq n-2 \quad (6.6.13).$$

Слід зауважити, що використовувати методи згладжування необхідно дуже обережно. У деяких випадках різкі зміни значень характеризують відповідні якісні зміни, які відбуваються в дослідному об'єкті. Тому при

згладжуванні таких експериментальних результатів ми можемо отримати неправдиву інформацію про такі явища.

Задача. Провести згладжування масиву температур (Т), вимір яких проводили через рівні проміжки часу:

№	Y _{експ.}	YR(пр.)	YR(3-х)	YR(5-ти)
1	0,9	0,9	0,97	0,992
2	2,12	2,036	1,98	1,995
3	2,92	3,003	3,06	2,998
4	4,15	4,049	3,99	4,038
5	4,9	4,993	5,05	4,998
6	6,1	6,031	5,97	6,044
7	6,92	7,009	7,06	7,024
8	8,15	8,057	8,04	8,004
9	9,05	9,02	9,00	8,957
10	9,8	9,80	9,83	9,910

Примітка. Y – експериментальні результати; YR – відповідні згладжені дані.

6.7. Побудова емпіричної лінії регресії

Багато залежностей між величинами, що характеризують хіміко-технологічні процеси, носять нелінійний характер (наприклад константа швидкості реакції експоненціально зростає із збільшенням температури). З метою визначення вигляду залежності між експериментальними величинами необхідно на основі результатів дослідів побудувати емпіричну лінію регресії (рис.6.7.1). Для цього весь діапазон зміни величини x розбивають на k рівних інтервалів довжиною Δx . Усі точки, що потрапили в інтервал із номером j , відносять до його середини $\bar{x}_j$. Для кожного інтервалу j обчислюють середнє арифметичне значення $\bar{y}_j$:

$$\bar{y}_j = \frac{1}{n} \sum_{i=1}^n y_{ji} \quad (6.7.1),$$

де n – кількість точок в інтервалі з номером j .

Потім послідовно з'єднують точки з координатами $(\bar{x}_j, \bar{y}_j)$ відрізками прямих ліній. Одержана ломана лінія є шуканою емпіричною лінією регресії. За виглядом цієї лінії можна підібрати відповідне рівняння регресії, що описує залежність між експериментальними величинами. Задачу визначення коефіцієнтів даного рівняння регресії вирішують із застосуванням методу найменших квадратів (шляхом варіювання коефіцієнтів підбираються такі його значення, що мінімізують суму квадратів відхилень розрахункових значень від експериментальних).

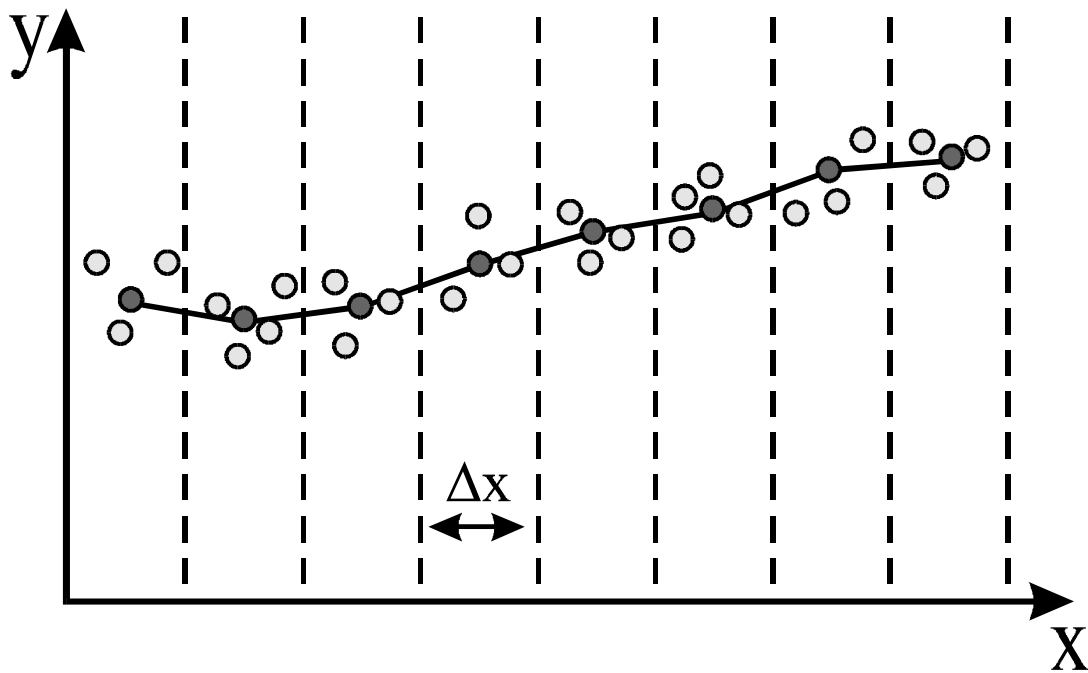


Рис.6.7.1.Емпірична лінія регресії

Якість рівняння регресії оцінюють за його здатності до прогнозування. Рівняння добре прогнозує експериментальні дані, тобто рівняння адекватно описує результати експерименту, якщо розбіжності між розрахунковими й експериментальними значеннями знаходяться в допустимих межах. Існують декілька способів перевірки адекватності рівняння регресії. У більшості випадків для цих цілей можна використовувати значення кореляційного відношення η або середньої відносної похибки прогнозування ε (%), що розраховуються за формулами:

$$\eta = \sqrt{1 - \frac{\sum_{i=1}^n (y_{e,i} - y_{p,i})^2}{\sum_{i=1}^n (y_{e,i} - y_c)^2}} \quad (6.7.2),$$

$$\varepsilon = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_{e,i} - y_{p,i}}{y_{e,i}} \right| \quad (6.7.3),$$

де y_e , y_p і y_c – відповідно експериментальні, розрахункові та середнє значення величини y ; n – кількість експериментальних значень y , що використовуються для перевірки адекватності.