ЛЕКЦІЯ №6
Регресія. Лінійні відображення. Інтерполяція.

Лінійні відображення.

1. В лінійному просторі L задана векторна функція **A** векторного аргумента $\vec{x}$, якщо кожному вектору $\vec{x}$ цього простору поставлений у відповідність деякий вектор $\vec{u} = A(\vec{x})$ цього ж простору. Векторна функція **A** називається лінійною, якщо вона має наступні властивості:

$A(\vec{x} + \vec{y}) = A(\vec{x}) + A(\vec{y})$, $A(\alpha\vec{x}) = \alpha A(\vec{x})$, де $\vec{x}$ і $\vec{y}$ – два довільних вектори простору L і α – довільне дійсне число. Лінійну вектор – функцію називають також лінійним перетворенням простору L, або **лінійним оператором** в цьому просторі. В подальшому, будемо при визначенні лінійної вектор – функції ігнорувати дужки аргументу. Тобто представляти вектор – функцію в наступному вигляді:

$$\vec{u} = A\vec{x} \qquad (1)$$

Розглянемо двомірний випадок відображення площини P саму в себе. Тобто $P \to \overline{P}$. Вважатимемо, що на площині P введено декартову систему координат.

Якщо точка $M(x_1;x_2)$ при відображенні переходить в точку $\overline{M}(y_1;y_2)$ то існують функції $y_1 = \varphi_1(x_1;x_2)$, $y_2 = \varphi_2(x_1;x_2)$. Якщо зазначені функції є лінійними відносно своїх аргументів, то відображення є **лінійним**.

Таким чином при лінійному відображені координати перетворюються за таким законом:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + b_1 \\ y_2 = a_{21}x_1 + a_{22}x_2 + b_2 \end{cases} \qquad (2)$$

a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2 – деякі постійні числа. В подальшому під поняттям лінійне відображення, будемо розуміти відображення при якому початок введеної системи координат знаходиться в точці $\overline{O}(b_1; b_2)$. Тоді лінійне відображення матиме вигляд:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 \\ y_2 = a_{21}x_1 + a_{22}x_2 \end{cases} \qquad (3)$$

Тоді матриця $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ називається матрицею відображення.

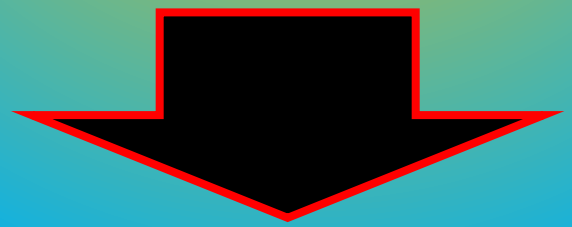
Відповідно $\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ називається визначником матриці А.

ГЕОМЕТРИЧНИЙ ЗМІСТ ВИЗНАЧНИКА МАТРИЦІ ОПЕРАТОРА

Виходячи з вищевикладеного можна зробити висновок, що $\det A$ характеризує коефіцієнт зміни площі фігур при відображенні. Так, наприклад квадрат з вершинами O(0; 0), A(1; 0), B(1; 1), C(0; 1) при лінійному відображені з матрицею A, перейде в паралелограм побудований на векторах $\overrightarrow{OA} = a_{11}\vec{i} + a_{21}\vec{j}$, $\overrightarrow{OC} = a_{12}\vec{i} + a_{22}\vec{j}$.

У випадку тривимірного простору, матриця відображення матиме вигляд:

$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Якщо $\det A = 0$, то відображення називається виродженим.



ВЛАСНІ ВЕКТОРИ І ВЛАСНІ ЗНАЧЕННЯ МАТРИЦІ ОПЕРАТОРА

2. При лінійному відображені існують вектори лінійного простору L які під впливом лінійного оператора **A** не відображаються в інші вектори, а лише «деформуються» уздовж свого напрямку (тобто залишаються паралельними самі собі при відображені). Такі вектори називаються **власними векторами оператора A**.

Розглянемо питання знаходження власних векторів оператора A в заданому базисі для тривимірного простору.

При визначеному базисі, координати вектора перетворюються наступним чином: $\vec{y} = A\vec{x}$. Для того, щоб вектор $\vec{x}$ був власним вектором оператора A, повинна виконуватись умова $\vec{y} = \lambda\vec{x}$ тобто:

$$A\vec{x} = \lambda\vec{x} \qquad (4)$$

Відповідно числа λ при яких виконується умова (4) називаються **власними значеннями** оператора A.

Визначимо в загальному вигляді власні вектори та власні числа лінійного оператора A для тривимірного простору. Розгорнута умова (4) матиме вигляд:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = \lambda x_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = \lambda x_3 \end{cases} \quad (5)$$

Після перетворень, отримаємо:

$$\begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 + a_{13}x_3 = 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + a_{23}x_3 = 0 \\ a_{31}x_1 + a_{32}x_2 + (a_{33} - \lambda)x_3 = 0 \end{cases} \quad (6)$$



Таким чином, для знаходження власних векторів маємо систему лінійних алгебраїчних рівнянь (6). Щоб дана система мала розв?язок, необхідно і достатньо рівності нулю визначника системи (6), тобто:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} = 0 . \qquad (7)$$

Або в такій формі:

$$\det(A-\lambda I) = 0 , \qquad (8)$$

де $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ – одинична матриця.

РЕГРЕСІЙНИЙ АНАЛІЗ. ЗАДАЧА ПРОГНОЗУВАННЯ

Рівняння (7) і (8) називається **характеристичним рівнянням матриці А**.

Визначивши характеристичні (власні) числа λ, знайдемо відповідно власні вектори $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ оператора А.

Регресійний аналіз. МНК.

Постановка задачі. Нехай потрібно вивчити (дослідити) залежність між фізичними величинами y та x, які зв'язані деякою функціональною залежністю $y = f(x)$, вид якої невідомий і його потрібно визначити на основі експерименту. Результатами експерименту є обчислені або виміряні значення $f(x)$ на сітці $x = \{x_0, x_1, \ldots, x_{n-1}\}$.

МЕТОД НАЙМЕНШИХ КВАДРАТІВ (МНК)

Оскільки експериментальні дані мають похибки (помилки), то задача ставиться таким чином: так обробити ці дані, щоб залежність між y та x відображалась з найбільшою точністю, щоб незакономірні помилки (похибки), які пов'язані з експериментом, були максимально згладжені.

Часто МНК називають **методом згладжування**. Він широко використовується в системах фільтрової обробки інформації, в системах екстраполяції (прогнозування) та ін.

Задача зводиться до визначення рівняння лінії – яка буде виражати реальну (потрібну) залежність. Вигляд лінії виражає вибір апроксимуючого багаточлена:

- лінійний $P(x) = a_0 + a_1 x$;
- степеневий $P(x) = a_0 a_1 x$;
- параболічний $P(x) = a_0 + a_1 x + a_2 x_2$;
- гіперболічний $p_x = a_0 + a_1$ тощо.

Нехай це буде лінійна тенденція

$$y = a_0 + a_1 x . \qquad (1)$$

Підстановка в рівняння (1) замість значень $x_0, x_1, \ldots, x_{n-1}$ дає нам ординати точок прямої, які не збігаються з ординатами точок $D_0, D_1, \ldots$. Тому різниці між ними

$$\Delta_0 = a_0 + a_j x_0 - y_0 = D_0 D_0;$$

$$\Delta_1 = a_0 + a_1 x_1 - y_1 = D_1 D_1;$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\Delta_{n-1} = a_0 + a_1 x_{n-1} - y_{n-1} = D_{n-1} D_{n-1}.$$

Як правило $\Delta_i \neq 0$.

Змінюючи параметри a_0 і a_1 (інакше, вибираючи то одну пряму то другу), можна змінювати величини відхилень $\Delta_0, \Delta_1, \Delta_{n-1}$, а потім знаходити мінімум суми відхилень (або нев'язок) і визначати коефіцієнти a_0, a_1; тобто

$$\sum_{i=0}^{n-1} (P_i(x) - y_i) \Rightarrow \min. \qquad (2)$$

Однак, даний спосіб супроводжується суттєвим недоліком. Різниці $\Delta_i = P_i(x) - y_i$ можуть виникати з різними знаками (+ або –).

Тому якщо знаходити суму відхилень у деяких експериментах, то можна отримати малу величину відхилення помилково, за рахунок взаємовилучення складових більшої величини, але різних знаків (рис. 1).

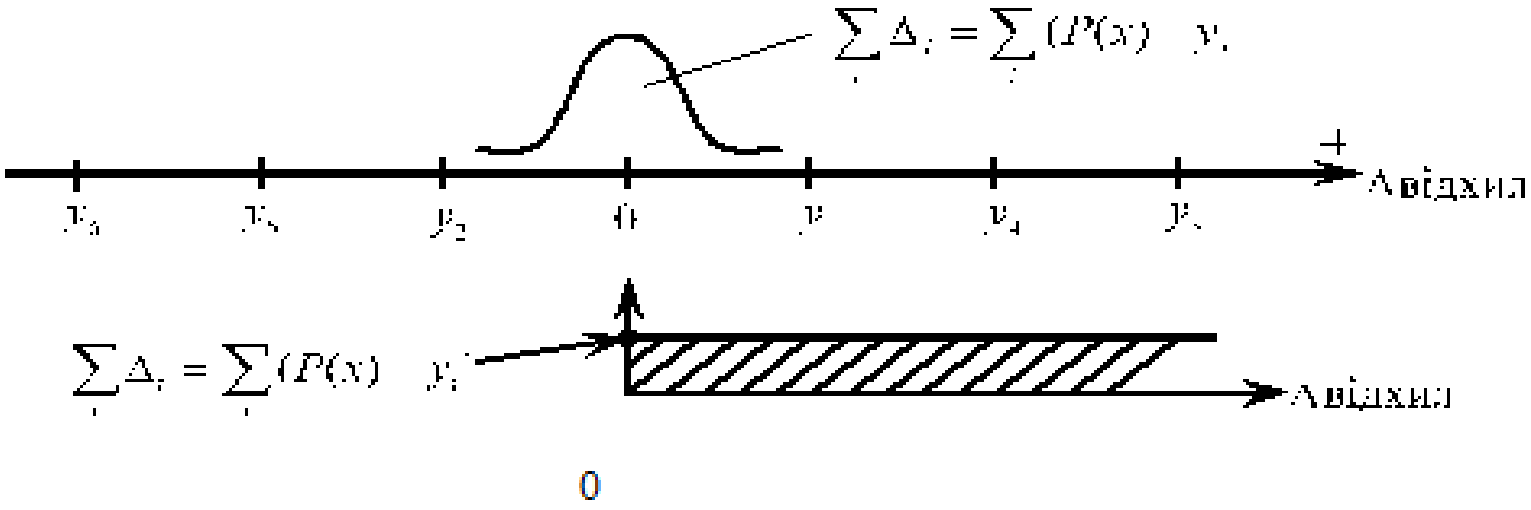


Рис. 1

Уникнути цього недоліку дозволяє міра відхилення (критерій), яка була запропонована французьким математиком Лежандром (і паралельно Гауссом) у 1806 році, – брати суму квадратів відхилень.

Таким чином, метод апроксимації, суть якого є мінімізація суми квадратів нев'язок ($P(x) - y_i$), веде свій початок від праць таких класиків математичної науки, як Лежандр і Гаусс, і називається **методом найменших квадратів**

$$S = \sum_{i=0}^{n-1} \left(P(x_i) - y_i \right)^2 \Rightarrow \min. \qquad (3)$$

Формула читається так: мінімум суми квадратів "нев'язок".

Одне з можливих узагальнень формули (3) – це "ваговий" МНК:

$$S_q = \sum_{i=0}^{n-1} q_i \left(P(x_i) - y_i \right)^2 \Rightarrow \min, \qquad (4)$$

де $q_i \equiv q_0, q_1, \ldots, q_{n-1}$ – дійсні додатні числа (ваговий коефіцієнт). У матричній формі: $\underline{Q} = diag\left(q_0, q_1, \ldots, q_{n-1}\right)$ – матриця вагових коефіцієнтів.

Причина (джерело) появи вагових коефіцієнтів – це неоднакова точність вимірювань, які проводяться в різних точках сітки X. Тому вибирають значення q тим більше, чим точніше проводилось вимірювання величини y_i.

Для простоти приймемо, що $q_0 = q_1 = \ldots = q_{n-1} = 1$.

Подамо формулу (3) у вигляді:

$$S = \sum_{i=0}^{n-1} \left[\sum_{k=0}^{m-1} a_k \varphi_x(x_i) - y_i \right]^2 \Rightarrow \min . \qquad (5а)$$

Таким чином, побудова оптимального апроксимуючого багаточлена $P(x)$ зводиться до знаходження коефіцієнтів, які мінімізують функцію $S \Rightarrow \min$, тобто

$$a_k = \arg\min S(a_k)$$
$$a_k \in R^m \qquad (6)$$

(аргумент, що мінімізує функцію $S(a_k)$), $R^m - m$ - мірний простір.

Як відомо з курсу математичного аналізу, необхідною умовою екстремуму (функції декількох змінних) є обернення в нуль її частинних похідних у цих змінних:

$$\frac{\partial S}{\partial a_k} = 0, \quad (k = 0,1,\ldots,m-1) \qquad (7)$$

Розглянемо задачу обчислення коефіцієнтів $a_к$ для полінома першого степеня (лінійна функція $\varphi_k(\alpha)$): $P(x_i) = a_0 + a_1 x_i$.

Підставляємо значення $P(x)$ у формулу (5а):

$$S = \sum_{i=0}^{n-1} (a_0 + a_1 x_i - y_i)^2 \qquad (8)$$

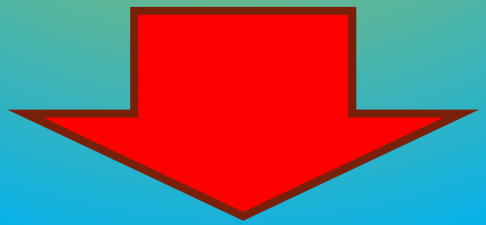
Запишемо систему частинних похідних і прирівняємо їх до нуля:

$$\left.\begin{aligned} \frac{\partial S}{\partial a_0} &= 0 \\ \frac{\partial S}{\partial a_1} &= 0 \end{aligned}\right\} \qquad (9)$$

Розкриваємо дужки у формулі (9.8):

$$[(a_0 + a_1 x_i) - y_i]^2 = (a_0 + a_1 x_i)^2 - 2(a_0 + a_1 x_i) y_i + y_i^2 =$$

$$= a_0^2 + 2a_0 a_1 x_i + a_1^2 x_i^2 - 2a_0 y_i - 2a_i x_i y_i + y_i^2.$$



Підставимо у систему (9):

$$\frac{\partial S}{\partial a_0} = \frac{\partial \sum_{i=0}^{n-1}}{\partial a_0}\left(a_0^2 + 2a_0a_ix_i + a_1^2x_i^2 - 2a_0y_i - 2a_1x_iy_i + y_i^2\right) = 0;$$

$$\sum_{i=0}^{n-1}\left(2a_0 + 2a_1x_i - 2y_i\right) = 0;$$

$$\sum_{i=0}^{n-1}a_0 + \sum_{i=0}^{n-1}a_1x_i - \sum_{i=0}^{n-1}y_i = 0;$$

$$na_0 + a_1\sum_{i=0}^{n-1}x_i = \sum_{i=0}^{n-1}y_i\ .$$

$$\frac{\partial S}{\partial a_1} = \frac{\partial}{\partial a_1}\sum_{i=0}^{n-1}\left(a_0^2 + 2a_0a_1x_i + a_1^2x_i^2 - 2a_0y_i - 2a_1x_iy_i + y_i^2\right) = 0;$$

$$\sum_{i=0}^{n-1}\left(2a_0x_i + 2a_1x_i^2 - 2x_iy_i\right) = 0;$$

$$\sum_{i=0}^{n-1}a_0x_i + \sum_{i=0}^{n-1}a_1x_i^2 - \sum_{i=0}^{n-1}x_iy_i = 0;$$

$$a_0\sum_{i=0}^{n-1}x_i + a_1\sum_{i=0}^{n-1}x_i^2 = \sum_{i=0}^{n-1}x_iy_i\ .$$

У результаті отримуємо систему рівнянь:

$$\left\{\begin{array}{l} na_0 + a_1\sum_{i=0}^{n-1} x_i = \sum_{i=0}^{n-1} y_i \\ a_0\sum_{i=0}^{n-1} x_i + a_1\sum_{i=0}^{n-1} x_i^2 = \sum_{i=0}^{n-1} x_i y_i \end{array}\right\} \qquad (10)$$

Для зручності можна ввести позначення

$$r_0 = \sum_{i=0}^{n-1} y_i; \quad r_1 = \sum_{i=0}^{n-1} x_i y_i$$

$$\delta_1 = \sum_{i=0}^{n-1} x_i; \quad \delta_2 = \sum_{i=0}^{n-1} x_i^2;$$

Розв'язуючи систему (10) відносно a_0 та a_1, маємо

$$\left\{\begin{array}{l} a_0 = \dfrac{r_0\delta_2 - r_1\delta_1}{n\delta_2 - \delta_1^2} \\ a_1 = \dfrac{n\delta_1 - r_0\delta_1}{n\delta_2 - \delta_1^2} \end{array}\right\} \qquad (11)$$

або після підстановки

$$\begin{cases} a_0 = \dfrac{\sum_{i=0}^{n-1} y_i \sum_{i=0}^{n-1} x_i^2 - \sum_{i=0}^{n-1} x_i y_i \sum_{i=0}^{n-1} x_i}{n\sum_{i=0}^{n-1} x_i^2 - \left(\sum_{i=0}^{n-1} x_i\right)^2} \\ a_1 = \dfrac{n\sum_{i=0}^{n-1} x_i y_i - \sum_{i=0}^{n-1} y_i \sum_{i=0}^{n-1} x_i}{n\sum_{i=0}^{n-1} x_i^2 - \left(\sum_{i=0}^{n-1} x_i\right)^2} \end{cases} \qquad (12)$$

Після обчислення коефіцієнти a_0 і a_1 отримують конкретні числові значення, які підставляються в апроксимуючий багаточлен $P(x)$.

Таким чином, методика оцінки результатів експериментів за допомогою МНК складається з етапів:

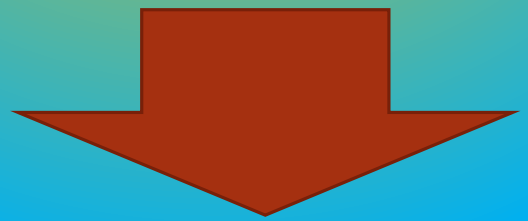
I. Обирають апроксимуючу функцію $\varphi(x)$.

II. Визначають $\min S$, як $\frac{\partial S}{\partial a_k} = 0$; тобто знаходять систему частинних похідних і прирівнюють їх до нуля.

III. Розв'язують систему рівнянь відносно коефіцієнтів a_k.

IV. Записують апроксимуючий поліном $P(x)$ з урахуванням числових значень коефіцієнтів $P(x)$.

Кількість лінійних рівнянь повинна бути не меншою за кількість незалежних коефіцієнтів. При цьому чим більше вимірювань, тобто чим більшою мірою система перевизначена (надлишок інформації), тим краще, бо тоді випадкові помилки (похибки) окремих вимірювань вилучають одна одну і рішення стає більш достовірним, тобто багаточлен P(x) більш адекватно описує систему або процес, що вивчається.



Приклад:

Підібрати апроксимуючий поліном другого степеню $y = a_0 + a_1 x + a_2 x^2$ для даних наведених в таблиці 1:

Таблиця 1

x	0,78	1,56	2,34	3,12	3,81
y	2,50	1,20	1,12	2,25	4,28

Розв'язок:

Використовуючи (10), отримаємо систему алгебраїчних рівнянь:

$$\begin{cases} 5a_0 + 11{,}61a_1 + 32{,}768a_2 = 11{,}350 \\ 11{,}61a_0 + 32{,}768a_1 + 102{,}761a_2 = 29{,}770 \\ 32{,}768a_0 + 102{,}761a_1 + 341{,}750a_2 = 94{,}604. \end{cases} \qquad (13)$$

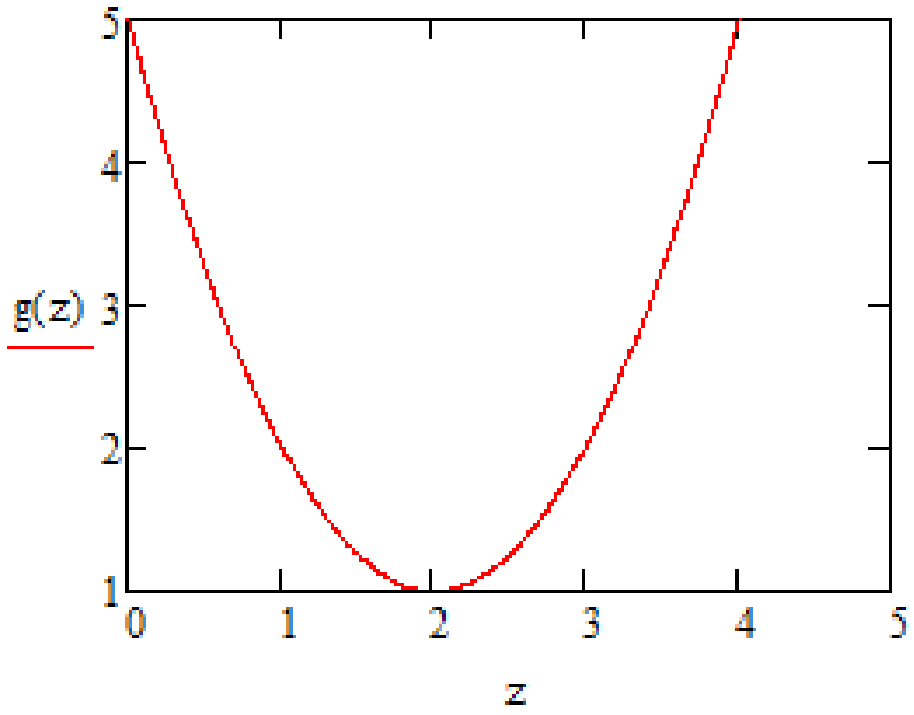
Розв'язавши систему (13), використовуючи, наприклад засоби Mathcad, або Exel, отримаємо значення відповідних коефіцієнтів:

$a_0 = 5{,}045; \quad a_1 = -4{,}043; \quad a_2 = 1{,}009.$

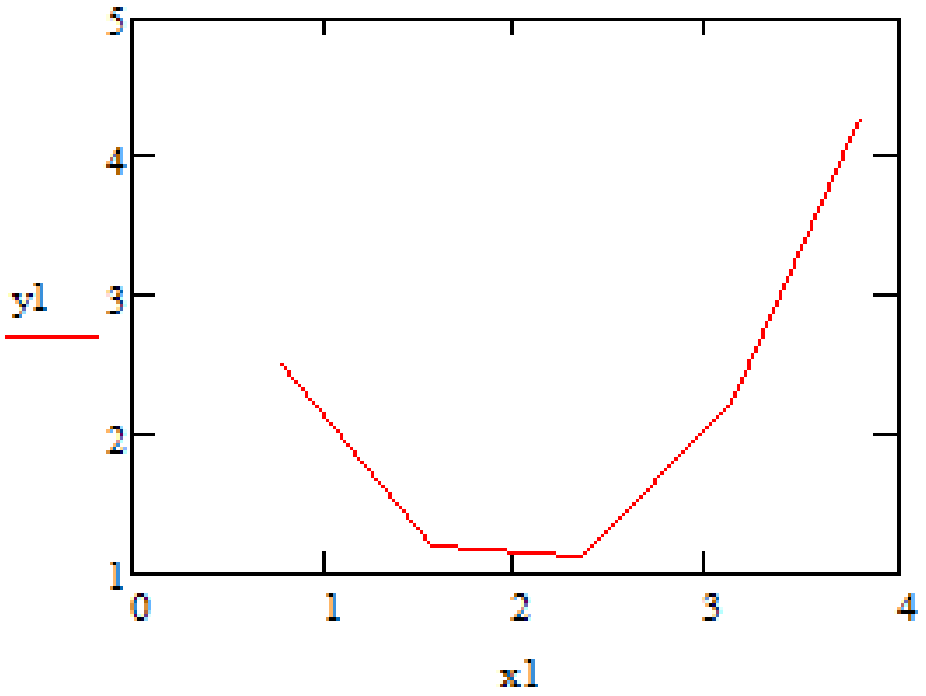
Тоді, шуканий поліном має вигляд:

$y = 5{,}045 - 4{,}043x + 1{,}009x^2.$

Графік отриманого поліному та розподілу статистичних даних з таблиці 1 наведено на рис. 2 а, б відповідно.



а)



б)

Рис. 2. Графік отриманого поліному та розподілу статистичних даних з таблиці 1. *а)*- поліном, *б)* - статистичні дані.

ЗАДАЧА ІНТЕРПОЛЯЦІЇ.

Інтерполяція.

Нехай на деякій множині задана система функцій $\varphi_0(x)$, $\varphi_1(x)$, ..., $\varphi_m(x)$, які в подальшому будемо вважати досить гладкими (наприклад, безперервно диференціюються) функціями. Назвемо цю систему основною.

Функції виду:

$$Q_m(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_m\varphi_m(x), \qquad (1)$$

де c_0, c_1, ..., c_m - постійні коефіцієнти, називаються узагальненими многочленами порядку m. Зокрема, якщо основна система складається з цілих невід'ємних степенів змінної x, т. е. $\varphi_0(x) = 1$, $\varphi_1(x) = x$, ..., $\varphi_m(x) = x_m$, то

$$Q_m(x) = c_0 + c_1x + \ldots + c_mx^m \qquad (2)$$

є звичайний поліном ступеня m.

Якщо

$$\varphi_0(x) = 1,$$

$$\varphi_1(x) = \cos x,$$

$$\varphi_2(x) = \sin x,$$

...

$$\varphi_{2m-1}(x) = \cos mx, \qquad (3)$$

$$\varphi_{2m}(x) = \sin mx,$$

то

$$Q_m(x) = a_0 + a_1 \cos x + b_1 \sin x \ldots + a_m \cos mx + b_m \sin mx \qquad (4)$$

називається тригонометричним поліномом (або тригонометричним многочленом) порядку m.

Завдання про наближення функцій ставиться таким чином: цю функцію f (x) потрібно замінити узагальненим многочленом Qm (x) заданого порядку m так, щоб відхилення (в сенсі σ або (Δi, Δh)) функції f (x) від узагальненого многочлена Qm (x) на зазначеній множині {x} було найменшим. При цьому многочлен Qm (x) в загальному випадку називається апроксимується.

Якщо множина {x} складається з окремих точок $x_0, x_1, ..., x_n$, то наближення називається дискретним. Якщо ж {x} є відрізок $a ? x ? b$, то наближення називається інтегральним.

На практиці часто користуються приближеннями функцій звичайним і тригонометричним поліномами.

В теорії дискретного наближення функцій має місце задача інтерполяції функцій. У разі звичайного полінома завдання інтерполяції формулюється в такий спосіб: для даної функції f(x) знайти поліном $Q_m(x)$ можливо нижчого ступеня m, що приймає в заданих точках xi ($i = 0, 1, 2, ..., n$; $x_i ? x_j$ при $i ? j$) ті ж значення, що і f (x), т. е. такий, що $Q_m(x_i) = f(x_i)$ ($i = 0, 1, 2, ..., n$). Такий поліном називають інтерполяційним, а точки x_i ($i = 0, 1, 2, ..., n$) - вузлами інтерполяції.

Нехай маємо функцію $f(x)$. Розглянемо відрізок $[a,b]$ на якому визначено дану функцію. Зафіксуємо на цьому відрізку $n+1$ значень аргумента x_i. При цьому покладемо

$$f(x_i) = P(x_i), \tag{5}$$

де $P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m$. Підставивши даний вираз для $P(x)$ в (1), отримаємо систему з $n+1$ рівнянь з $m+1$ невідомими коефіцієнтами $a_0, a_1 \ldots a_m$:

$$\begin{aligned} &a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_m x_1^m = f(x_1) \\ &\text{----------} \\ &a_0 + a_1 x_{n+1} + a_2 x_{n+1}^2 + \ldots + a_m x_{n+1}^m = f(x_{n+1}). \end{aligned} \tag{6}$$

Визначник складений з $x_1, x_2 \ldots x_n$ має назву визначник Вандермонда:

$$W(x_1, x_2, \ldots x_n) = \begin{vmatrix} 1 & x_1 \ldots x_1^k \\ 1 & x_2 \ldots x_2^k \\ \ldots & \ldots \\ 1 & x_k \ldots x_k^k \end{vmatrix} \tag{7}$$

Для визначення $P(x) = a_0 + a_1 x + a_2 x^2 + ... + a_m x^m$ необхідно розв'язати систему рівнянь (2). Але існує більш раціональний шлях.

Розглянемо багаточлен $P_i(x) = \dfrac{(x - x_1)....(x - x_{i-1})(x - x_{i+1}).....(x - x_{n+1})}{(x_i - x_1)....(x_i - x_{i-1})(x_i - x_{i+1}).....(x_i - x_{n+1})}$,

$i = 1,2,....n+1$. При $x = x_i$, $P_i(x) = 1$, при $x = x_j \quad i \neq j$, $P_i(x_j) = 0$

Тому ми можемо представити $P(x) = a_0 + a_1 x + a_2 x^2 + ... + a_m x^m$ у наступному вигляді:

$$P(x) = \sum_{i=1}^{n+1} f(x_i) P_i(x) \qquad (8)$$

Окрім інтерполяційного многочлена Лагранжа (8) використовується також метод розділених різниць з отриманням інтерполяційного многочлена Ньютона.