Заняття 9

Метод найменших квадратів

В останній презентації ми познайомимося з найбільш відомим додатком **ФДЗ**, яке знаходить саме широке застосування в різних галузях науки і практичної діяльності. Ви навчитеся вирішувати задачі **методом найменших квадратів**. Спочатку загальна постановка задачі + приклад:

Нехай в деякій предметній області досліджуються показники X, Y, які мають кількісний вираз. При цьому є всі підстави вважати, що показник Y залежить від показника X. Це може бути як науковою гіпотезою, так і зосновуватися на елементарному здоровому глузді. Позначимо через:

X – торгову площу продовольчого магазину, кв.м.,
Y – річний товарообіг продовольчого магазину, млн. грн.
Зрозуміло, що чим більше площа магазину, тим в більшості випадків буде більшим його товарообіг.

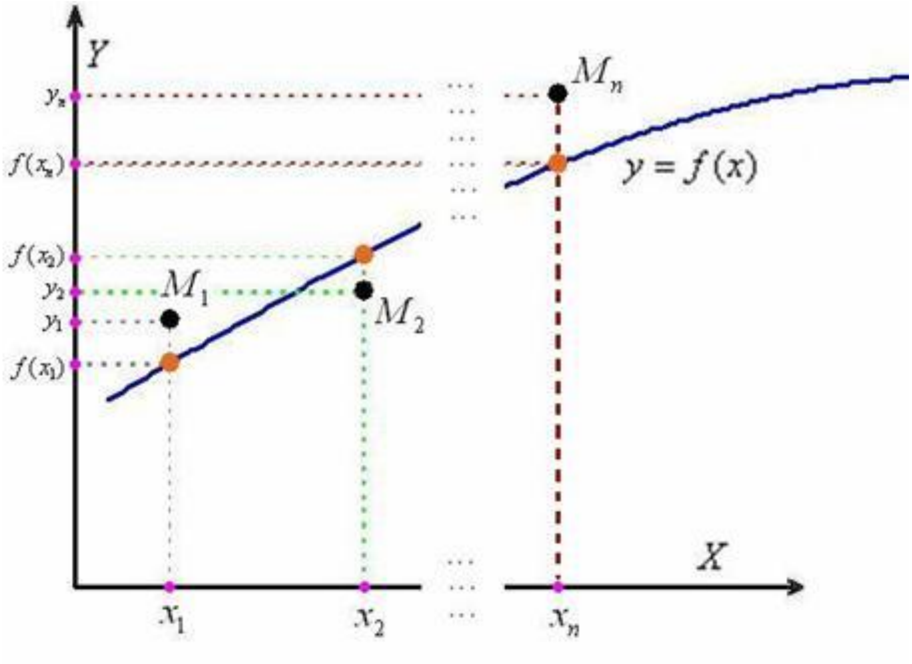
Уявімо, що після проведення n спостережень/опитів/підрахунків в нашому розпорядженні з`являються числові дані:

X	x_1	x_2	$\cdots$	x_n
Y	y_1	y_2	$\cdots$	y_n

С гастрономами, думаю, все зрозуміло: x_1 – це площа 1-го магазину, y_1 – його річний товарообіг, x_2 – площа 2-го магазину, y_2 – його річний товарообіг і т.д.
Табличні дані також можна записати у вигляді точок $M_1(x_1; y_1), M_2(x_2; y_2), \ldots, M_n(x_n; y_n)$ і зобразити в звичній для нас **декартовій системі** XOY.

Відповімо на важливе питання: **скільки точок потрібно для якісного дослідження?** Чим більше, тим краще. Мінімально припустимий набір складає 5-6 точок. Окрім того, при невеликій кількості даних у вибірку не можна включати «аномальні» результати. Так, наприклад, невеликий елітний магазин може виручати на порядки більше за «своїх колег», викривляючи тим самим загальну закономірність, яку і треба знайти!
Якщо зовсім просто – нам необхідно підібрати функцію $y = f(x)$, **графік** якої проходить як можна ближче до точок $M_1, M_2, \ldots, M_n$. Таку функцію називають ***апроксимуючою*** *(апроксимація –наближення)* або ***теоретичною функцією***. Взагалі то кажучи, тут одразу з`являється очевидний «претендент» – многочлен високої степені, графік якого проходить через ВСІ точки. Але цей варіант складний, а іноді і взагалі просто некоректний *(так як графік буде весь час «петляти»і погано відображати головну тенденцію)*.

Таким чином, розшукувана функція повинна бути достатньо простою і в той же час відображати залежність адекватно. Один з методів знаходження таких функцій і називається **методом найменших квадратів**. Спочатку розберемо його сенс в загальному вигляді. Нехай деяка функція $y = f(x)$ наближає експериментальні дані $M_1(x_1; y_1), M_2(x_2; y_2), \ldots, M_n(x_n; y_n)$:



Як оцінити точність даного наближення? Обчислимо $f(x_1), f(x_2), \ldots, f(x_n)$ і різниці (відхилення) $e_1 = y_1 - f(x_1), e_2 = y_2 - f(x_2), \ldots, e_n = y_n - f(x_n)$ між експериментальними і функціональними значеннями *(вивчаємо рисунок)*. Перша думка, яка приходить в голову – це оцінити, наскільки велика сума $e_1 + e_2 + \ldots + e_n$, але проблема полягає в тому, що різниці можуть бути і від`ємними *(наприклад,* $e_2 = y_2 - f(x_2) < 0$*)* і відхилення в результаті такого сумування будуть взаємознищуватися. Тому в якості оцінки точності наближення напрошується прийняти суму **модулів** відхилень:

$|e_1| + |e_2| + \ldots + |e_n|$ або в згорнутому вигляді: $\sum_{i=1}^{n} |e_i|$ *(якщо хтось не знає:* $\sum$ *– це значок суми, а* i *– допоміжна змінна - «лічильник», яка приймає значення від 1 до* n*).*

Наближуючи експериментальні точки різними функціями, ми будемо отримувати різні значення $\sum_{i=1}^{n} |e_i|$, і вочевидь, де ця сума менша – та функція і точніша.

Такий метод існує і називається він *методом найменших модулів*. Але на практиці більше розповсюдження отримав *метод найменших квадратів*, в якому можливі від`ємні значення ліквідуються не модулем, а піднесенням відхилень до квадрату:
$e_1^2 + e_2^2 + \ldots + e_n^2$, після чого зусилля направлені на підбір такої функції $y = f(x)$, щоб сума квадратів відхилень $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2$ була якомога меншою. Звідси і назва методу.
І зараз ми повертаємося до другого важливого моменту: як зазначалося вище, підбираєма функція повинна бути достатньо простою – але і таких функцій теж не мало: ***лінійна***, ***гіперболічна***, е***экспоненціальна***, ***логарифмічна***, ***квадратична*** і т.д. І, звісно ж, тут одразу б хотілося «скоротити поле діяльності». Який клас функцій обрати для дослідження? Примітивний, але ефективний прийом:

– Простіше за все зобразити точки $M_1, M_2, \ldots, M_n$ на рисунку і проаналізувати їх розташування. Якщо вони мають тенденцію розташовуватися по прямій, то слід шукати **рівняння прямої** $y = f(x) = ax + b$ з оптимальними значеннями a і b. Іншими словами, задача полягає в знаходженні ТАКИХ коефіцієнтів a, b – щоб сума квадратів відхилень $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$ була найменшою.

Якщо ж точки розташовані, наприклад, по **гіперболі**, то зрозуміло, що лінійна функція буде давати погане наближення. В цьому випадку шукаємо найбільш «вигідні» коефіцієнти a, b для рівняння гіперболи $y = f(x) = \frac{a}{x} + b$ – ті, які дають мінімальну суму квадратів $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{i=1}^{n} \left(y_i - \left(\frac{a}{x_i} + b\right)\right)^2$.

А тепер зверніть увагу, що в обох випадках річ іде про **функцію двох змінних**, аргументами якої є параметри розшукуваних залежностей:

$$F(a\,;b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

$$F(a\,;b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \left(\frac{a}{x_i} + b\right)\right)^2$$

І нам треба вирішити стандартну задачу – знайти **мінімум функції двох змінних**.

Згадаємо про наш приклад: припустимо, що «магазині» точки $M_1(x_1;\, y_1), M_2(x_2;\, y_2), \ldots, M_n(x_n;\, y_n)$ мають тенденцію розташовуватися по прямій лінії і є всі підстави вважати що є *лінійна залежність* $y = ax + b$ товарообігу від торгівельної площі. Знайдемо ТАКІ коефіцієнти «а» і «бе», щоб сума квадратів відхилень $F(a\,;b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$ була найменшою. Все як зазвичай – спочатку **частинні похідні 1-го порядку**. Згідно з **правилом лінійності** диференціювати можна прямо під значком суми:

$$\frac{\partial F}{\partial a} = \left(\sum_{i=1}^{n} (y_i - (ax_i + b))^2\right)'_a = \sum_{i=1}^{n} \left[2(y_i - (ax_i + b)) \cdot (y_i - (ax_i + b))'_a\right] =$$

$$= 2\sum_{i=1}^{n} \left[(y_i - ax_i - b) \cdot (0 - (x_i + 0))\right] = 2\sum_{i=1}^{n} \left[(y_i - ax_i - b) \cdot (-x_i)\right] = 2\sum_{i=1}^{n} (ax_i^2 + bx_i - x_i y_i)$$

$$\frac{\partial F}{\partial b} = \left(\sum_{i=1}^{n} (y_i - (ax_i + b))^2\right)'_b = \sum_{i=1}^{n} \left[2(y_i - (ax_i + b)) \cdot (y_i - (ax_i + b))'_b\right] =$$

$$= 2\sum_{i=1}^{n} \left[(y_i - ax_i - b) \cdot (0 - (0 + 1))\right] = 2\sum_{i=1}^{n} (ax_i + b - y_i)$$

Складемо стандартну систему:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 0 \\ \dfrac{\partial F}{\partial b} = 0 \end{cases} \Rightarrow \begin{cases} 2\sum\limits_{i=1}^{n} (ax_i^2 + bx_i - x_i y_i) = 0 \\ 2\sum\limits_{i=1}^{n} (ax_i + b - y_i) = 0 \end{cases}$$

Скорочуємо кожне рівняння на «двійку» і, крім того, «розвалюємо» суми:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0 \\ a\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} b - \sum_{i=1}^{n} y_i = 0 \end{cases} \Rightarrow \begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i = 0 \\ a\sum_{i=1}^{n} x_i + (\underbrace{b+b+\ldots+b}_{n\text{ раз}}) - \sum_{i=1}^{n} y_i = 0 \end{cases}$$

***Примітка**: самостійно проаналізуйте, чому «а» і «бе» можна винести за значок суми. До речі, формально це можна зробити і з сумою* $\sum_{i=1}^{n} b = b\sum_{i=1}^{n} 1 = b\cdot(\underbrace{1+1+\ldots+1}_{n\text{ раз}}) = bn$

Перепишемо систему в «прикладному» вигляді:

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

після чого починається прорисовуватися алгоритм вирішення нашоїй задачі:

Координати точок $M_1(x_1; y_1), M_2(x_2; y_2), \ldots, M_n(x_n; y_n)$ ми знаємо? Знаємо. Суми $\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} y_i, \sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i y_i$ знайти можемо? Легко. Складаємо найпростішу **систему двох лінійних рівнянь з двома невідомими** («а» і «бе»). Систему вирішуємо, наприклад, **методом Крамера**, в результаті чого отримуємо стаціонарну точку $S(a^*; b^*)$. Перевіряючи **достатні умови екстремуму**, можна впевнитися, що в даній точці функція $F(a;b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$ досягає саме **мінімуму**. Перевірка пов`язана з додатковими викладками і тому залишимо її за кадром *(при необхідності кадр, якого не вистачає, можна подивитися тут)*. Робимо остаточний висновок:

Функція $y = f(x) = a^* x + b^*$ найкращим чином наближає експериментальні точки $M_1, M_2, \ldots, M_n$. Грубо кажучи, її графік проходить максимально близько до цих точок. В традиціях е**конометрики** отриману апроксимуючу функцію також називають ***рівнянням парної лінійної регресії***.

Розглянута задача має велике практичне значення. В ситуації з нашим прикладом, рівняння $y = f(x) = a^* x + b^*$ дозволяє прогнозувати, який товарообіг *(«ігрек»)* буде у магазина при тому чи іншому значенні торгівельної площі *(тому чи іншому значенні «ікс»)*. Так, отриманий прогноз буде лише прогнозом, але в багатьох випадках він виявиться достатньо точним.

Я розберу всього лише одну задачу з «реальними» числами, оскільки ніяких труднощів в ній не має – всі обчислення на рівні шкільної програми 7-8 класу. В 95 відсотках випадків вам буде запропоновано відшукати як раз лінійну функцію, але в самому кінці презентації я покажу, що взагалі не важко відшукати рівняння оптимальної гіперболи, експоненти і деяких інших функцій.

Задача

В результаті дослідження взаємозв`язку двох показників, отримані наступні пари чисел:

x_i	1	2	3	4	5
y_i	5,3	6,3	4,8	3,8	3,3

Методом найменших квадратів знайти лінійну функцію, яка найкращим чином наближає емпіричні *(дослідні)* дані. Зробити рисунок, на якому в декартовій прямокутній системі координат побудувати експериментальні точки $M_i(x_i;y_i)$ і графік апроксимуючої функції $y=f(x)=ax+b$. Знайти суму квадратів відхилень між емпіричними y_i і теоретичними $f(x_i)$ значеннями. З`ясувати, чи буде функція $y=6{,}65e^{-0{,}15x}$ краще *(з точки зору методу найменших квадратів)* наближати експериментальні точки.

Зауважте, що «іксові» значення – натуральні, і це має характерний сенс, про який я розкажу трохи пізніше; але вони, відповідно, можуть бути і дробами. Окрім того, в залежності від змісту тієї або іншої задачі як «іксові», так і «ігрекові» значення повністю або частково можуть бути від`ємними. А в нас дана «своя» задача, і ми починаємо її вирішення:

Коефіцієнти a,b оптимальної функції $y=ax+b$ знайдемо як вирішення системи:

$$\begin{cases} a\sum x_i^2+b\sum x_i=\sum x_iy_i \\ a\sum x_i+bn=\sum y_i \end{cases}$$

З метою більш компактного запису змінну-«лічильник» можна опустити, оскільки і так зрозуміло, що сумування відбувається від 1 до $n=5$.

Розрахунок необхідних сум зручніше оформити в табличному вигляді:

x_i	1	2	3	4	5	$\sum x_i=$	15
y_i	5,3	6,3	4,8	3,8	3,3	$\sum y_i=$	23,5
x_i^2	1	4	9	16	25	$\sum x_i^2=$	55
x_iy_i	5,3	12,6	14,4	15,2	16,5	$\sum x_iy_i=$	64

Обчислення можна провести на мікрокалькуляторі,але набагато краще використовувати Excel – і скоріше, і без помилок:

Таким чином, отримуємо наступну **систему**:

$$\begin{cases} 55a+15b=64 \\ 15a+5b=23{,}5 \end{cases}$$

Тут можна помножити друге рівняння на 3 і **з 1-го рівняння почленно відняти 2-е**. Але це везіння – на практиці системи частіше не подарункові, і в таких випадках рятує **метод Крамера**:

$$\Delta=\begin{vmatrix} 55 & 15 \\ 15 & 5 \end{vmatrix}=55\cdot5-15\cdot15=275-225=50\neq0$$

, значить, система має єдине рішення.

$$\Delta_a = \begin{vmatrix} 64 & 15 \\ 23,5 & 5 \end{vmatrix} = 64 \cdot 5 - 23,5 \cdot 15 = 320 - 352,5 = -32,5$$

$$a = \frac{\Delta_a}{\Delta} = \frac{-32,5}{50} = -0,65$$

$$\Delta_b = \begin{vmatrix} 55 & 64 \\ 15 & 23,5 \end{vmatrix} = 55 \cdot 23,5 - 15 \cdot 64 = 1292,5 - 960 = 332,5$$

$$b = \frac{\Delta_b}{\Delta} = \frac{332,5}{50} = 6,65$$

Виконаємо перевірку. Розумію, що не хочеться, але навіщо ж пропускати помилки там, де їх можна стовідсотково не пропустити? Підставимо знайдене рішення $a = -0,65, \quad b = 6,65$ в ліву частину кожного рівняння системи:

$$55 \cdot (-0,65) + 15 \cdot 6,65 = -35,75 + 99,75 = 64$$

$$15 \cdot (-0,65) + 5 \cdot 6,65 = -9,75 + 33,25 = 23,5$$

Отримані праві частини відповідних рівнянь, значить, система вирішена правильно.

Таким чином, шукана апроксимуюча функція: $y = f(x) = -0,65x + 6,65$ – зі *всіх лінійних функцій* експериментальні дані найкращим чином наближає саме вона.

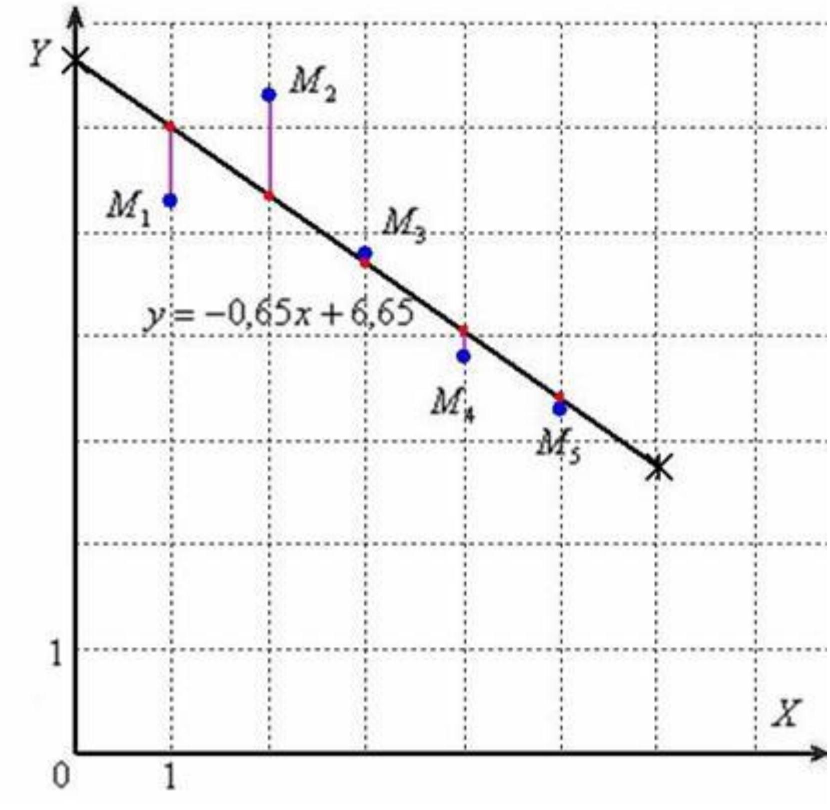
На відміну від ***прямої*** залежності товарообігу магазину від його площі, знайдена залежність є ***зворотною*** *(принцип «чим більше – тим менше»)*, і цей факт одразу виявляється по від`ємному **кутовому коефіцієнту**. Функція $y = -0,65x + 6,65$ сповіщає нам про те, що зі збільшення деякого пказника X на 1 одиницю значення залежного показника Y зменшується *в середньому* на 0,65 одиниць. Як кажуть, чим вища ціна на гречку, тим менше її продано.

Для побудови графіку апроксимуючої функції знайдемо два її значення:

$$f(0) = -0,65 \cdot 0 + 6,65 = 6,65$$

$$f(6) = -0,65 \cdot 6 + 6,65 = -3,9 + 6,65 = 2,75$$

і виконаємо рисунок:



Побудована пряма називається ***лінією тренду*** *(а саме – лінією лінійного тренду, тобто в загальному випадку тренд – це не обов`язково пряма лінія)*.

Обчислимо суму квадратів відхилень $\sum e_i^2 = \sum (y_i - f(x_i))^2$ між емпіричними y_i і теоретичними $f(x_i)$ значеннями. Геометрично – це сума квадратів довжин «малинових» відрізків *(два з яких настільки малі, що їх взагалі не видно)*.

Обчислення зведемо в таблицю:

x_i	1	2	3	4	5		
y_i	5,3	6,3	4,8	3,8	3,3		
$f(x_i)$	6	5,35	4,7	4,05	3,4		
$(y_i - f(x_i))^2$	0,49	0,9025	0,01	0,0625	0,01	$\sum e_i^2 =$	1,475

Їх можна знову ж таки провести вручну, на всяк випадок наведу приклад для 1-й точки:

$f(x_1) = f(1) = -0{,}65 \cdot 1 + 6{,}65 = 6$

$(y_1 - f(x_1))^2 = (5{,}3 - 6)^2 = (-0{,}7)^2 = 0{,}49$

але набагато ефективніше поступити вже відомим чином:

Ще раз повторимо: **в чому сенс отриманого результату?** Зі *всіх лінійних функцій* у функції $y = -0{,}65x + 6{,}65$ показник $\sum e_i^2$ є найменшим, тобто в своєму сімействі це найкраще наближення. І тут, до речі, не випадкове заключне питання задачі: а раптом запропонована експоненціальна функція $y = g(x) = 6{,}65e^{-0{,}15x}$ буде краще наближати експериментальні точки?

Знайдемо відповідну суму квадратів відхилень $\sum \varepsilon_i^2$ – щоб розрізняти, я позначу їх літерою «епсілон». Техніка точно така ж сама:

x_i	1	2	3	4	5		
y_i	5,3	6,3	4,8	3,8	3,3		
$g(x_i)$	5,72	4,93	4,24	3,65	3,14		
$(y_i - g(x_i))^2$	0,1795	1,8867	0,3133	0,0226	0,0252	$\sum \varepsilon_i^2 \approx$	2,4274

І знову на всяк пожежний обчислення для 1-й точки:

$g(x_1) = g(1) = 6{,}65e^{-0{,}15 \cdot 1} \approx 5{,}72$

$(y_1 - g(x_1))^2 \approx (5{,}3 - 5{,}72)^2 \approx 0{,}1795$

В Excel користуємося стандартною функцією *EXP (синтаксис можна подивитися в екселівській Довідці)*.

Висновок: $\sum \varepsilon_i^2 > \sum e_i^2$, значить, експоненціальна функція $y = 6{,}65e^{-0{,}15x}$ наближає експериментальні точки гірше, ніж пряма $y = -0{,}65x + 6{,}65$.

Але тут слід зазначити, що «гірше» – це **ще не означає**, що погано. Графік цієї експоненціальної функції теж проходить близько до точок M_1, M_2, M_3, M_4, M_5 – да так, що без аналітичного дослідження і сказати важко, яка з функцій точніша.

На цьому рішення закінчене.