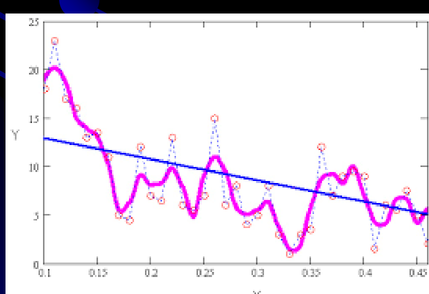
Лекція 4-5 (4 год., заняття 4-5).
Апроксимація функцій і процесів в системах

- 1. Модель «вхід - вихід»
- 2. Прикладні цілі та задачі апроксимації
- 3. Методи та алгоритми: метод найменших квадратів, «ковзне середнє», експоненціальне згладжування.
- 4. Інструментарій апроксимації даних в Excel та Mathcad.

«Хоча це може здатися парадоксом, вся наука підпорядкована ідеї апроксимації »

Б. Рассел



Апроксимація (<u>лат.</u> *approximatio* — наближення, proxima - найближча) науковий метод, суть якого полягає в заміні одних об'єктів іншими, у якомусь сенсі близькими до вихідних.

1. Модель «вхід - вихід»

При дослідженні процесів в системах, явищ різного походження, часто виникає задача вивчення та оцінювання взаємозалежностей (їх характеристик, показників, інформаційних потоків) з метою отримання нових достовірних даних для потреб моделювання, прогнозування, прийняття правильних рішень, нових планів, стратегій, знань.

Для розглядання такої задачі можна застосувати метод «Чорного ящика» (У.Р.Ешбі)

Система, процес, явище
$Y=f(x,a_к)$

На рис.1 показана загальна модель статистичних залежностей:
рис.1а – детермінованих,
рис.1б – стохастичних зв’язків.

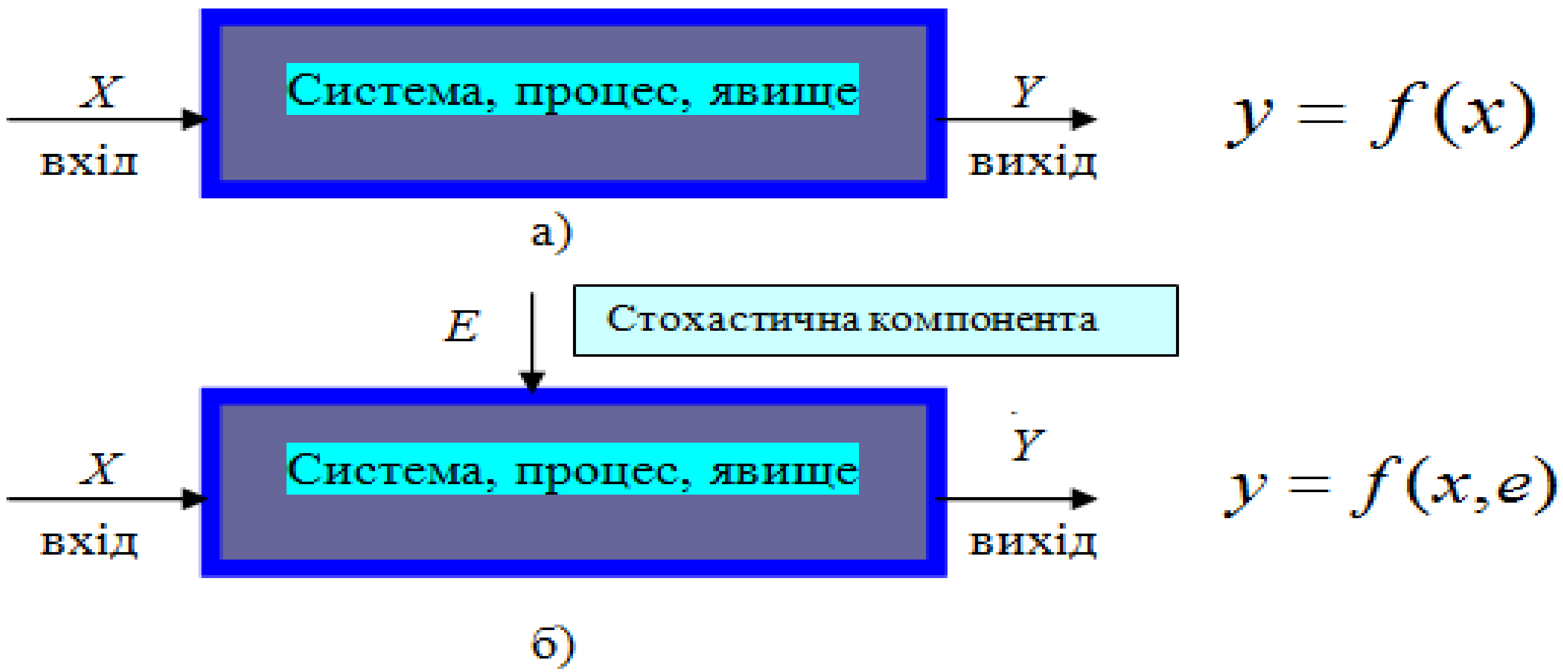


Рис. 1

Змінні X, Y, E визначають відповідно вхідний вплив (X), результат впливу (Y) та випадковий вплив (E) і подаються у вигляді векторних змінних різних розмірностей: $X = (x_1, x_2, \cdots, x_n)^T$, $Y = (y_1, y_2, \ldots, y_m)^T$, $E = (e_1, e_2, \ldots, e_k)^T$. Крім того, компоненти векторів X, Y, E можуть бути функціями від часу, тобто представляти собою часові процеси.

Змінна X описує умови функціонування об'єкта і в різних задачах називається: незалежною змінною, екзогенною, факторною ознакою, предиктором (передбачувачем), пояснюючою змінною, регресором.

Вихідна змінна Y характеризує поведінку, результат функціонування системи і в статистичних моделях її називають: залежною змінною, ендогенною, відгуком, результуючою, пояснюваною змінною, регресандом.

Вектор E (зовнішній, випадковий вплив) складається з латентних (прихованих) стохастичних компонентів, які відображають вплив неврахованих (невизначених, невідомих) факторів, а також випадкові помилки при вимірюванні аналізуємих показників.

В задачах статистичного аналізу компоненти векторів *X*, *Y*, *E* можуть бути різних типів, що суттєво впливає на вибір методів і моделей:

- **кількісні**, які можна вимірювати за визначеною шкалою (грошові показники, чисельності, розмірні характеристики систем, процесів, явищ);
- **порядкові**, що дозволяють упорядковувати дані за якісною ознакою степені прояву вивчаємої властивості (рівень освіти, професійності, умов існування, можливостей тощо);
- **класифікаційні**, що дозволяють розділяти сукупність об'єктів (даних) на однорідні за аналізуємою властивістю, які неможливо упорядкувати (види, класи, мотиви, професії тощо).

В процесі проведення математико–статистичного моделювання необхідно враховувати **типи взаємозв'язку змінних**: **рівноправний** (т.б. зв'язок можна розглядати як в один бік, так і в інший) або **нерівноправний** (причина – наслідок), а також **види даних**: **перехресні дані** (вибірки даних) та **часові дані** (часові ряди).

Перехресними є дані за якимось показником, що отримані для різних однотипних об’єктів (підприємств, областей, регіонів) в один момент часу (тобто часова приналежність несуттєва).

Часовий ряд характеризує один об’єкт в різні моменти часу.
Послідовні значення часових рядів можуть бути пов’язані між собою певними залежностями, що визначаються закономірними відхиленнями від загальної тенденції розвитку, або виявляються часові лаги (з англ., lag – запізнення).

База статистичних даних основується на спостереженнях.
Формуючи спостереження, слід забезпечити порівнянність даних у просторі та часі.
Тому початкові дані (факторні ознаки) повинні підпорядковуватись наступним вимогам:

вимоги

Вимоги до початкових даних

однаковий ступінь агрегування

однорідна структура одиниць сукупності (вибірки)

одні й ті самі методи розрахунку показників у часі чи просторі

однакова періодичність обліку окремих змінних

однакові зовнішні умови

кожний із факторів *Xi* має бути обґрунтований теоретично

вибирати доцільно лише значущі фактори *Xi*, що суттєво впливають на досліджуванні показники *Y*

кількість факторів не повинна перевищувати однієї третини числа спостережень у вибірці (довжини часового ряду)

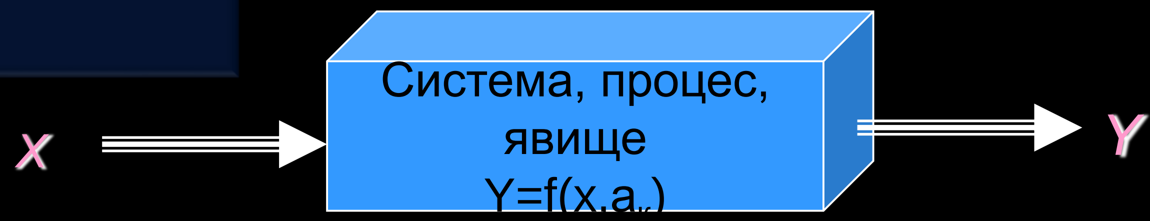
Крім того, потрібно враховувати та оцінювати можливість виникнення таких явищ, як мультиколінеарність та автокореляція

Ю. Бродський

Висновок:

Отже, в процесі проведення статистичного аналізу та математико-статистичного моделювання слід враховувати різні типи (компоненти векторів X, Y) та види (перехресні, часові) даних, типи взаємозв'язку (рівноправний, нерівноправний), види залежностей, зв'язків між явищами та процесами (функціональний, стохастичний, кореляційний), а також тип задачі дослідження: однофакторна $y = f(x)$ та багатофакторна $y = f(x_1, x_2, \ldots, x_n)$

2. Прикладні цілі та задачі апроксимації



● **<u>Загальну задачу можна сформулювати так:</u>**

на основі спостережень (вимірювань) змінних *X* та *Y* побудувати функцію, що найкраще буде відновлювати значення відгуку *Y*, відповідно заданим значенням предиктора *X*.

● Розв'язок даної задачі передбачає вибір математичного виразу для опису залежності та критерію якості апроксимації, відповідно до якого буде визначатись найкращий спосіб відновлення значень *Y*.

● Однак, перш ніж розв'язувати задачу потрібно визначити прикладну мету (ціль) проведення аналізу.

● Ясність прикладної цілі і відповідної задачі визначить послідовність виконання різних етапів проведення аналізу, вибір загальної структури функції, моделі, інтерпретацію отриманих результатів.

Ю. Бродський

Мета 1:

- *встановлення факту наявності (відсутності) статистично значимого зв'язку між змінними Y та X.*

Для даної задачі вибір виду функції f має другорядне значення і часто, навіть, не виникає питання побудови функції f. Задачі такого типу розв'язуються методами кореляційного аналізу, рангових кореляцій та за допомогою аналізу таблиць сполучення.

Мета 2:

- *відновлення (прогноз) значень результуючої змінної (відгуку) Y відповідно заданим значенням вхідної змінної (предиктора) X .*

Вибір функції апроксимації f також є підпорядкована мета, оскільки важливим є лише значення функції $f(X)$, а не її структура , тобто функція f повинна показати числову залежність змінних Y та X , а не їх змістовний зв'язок (фізичний, економічний тощо).

Ю. Бродський

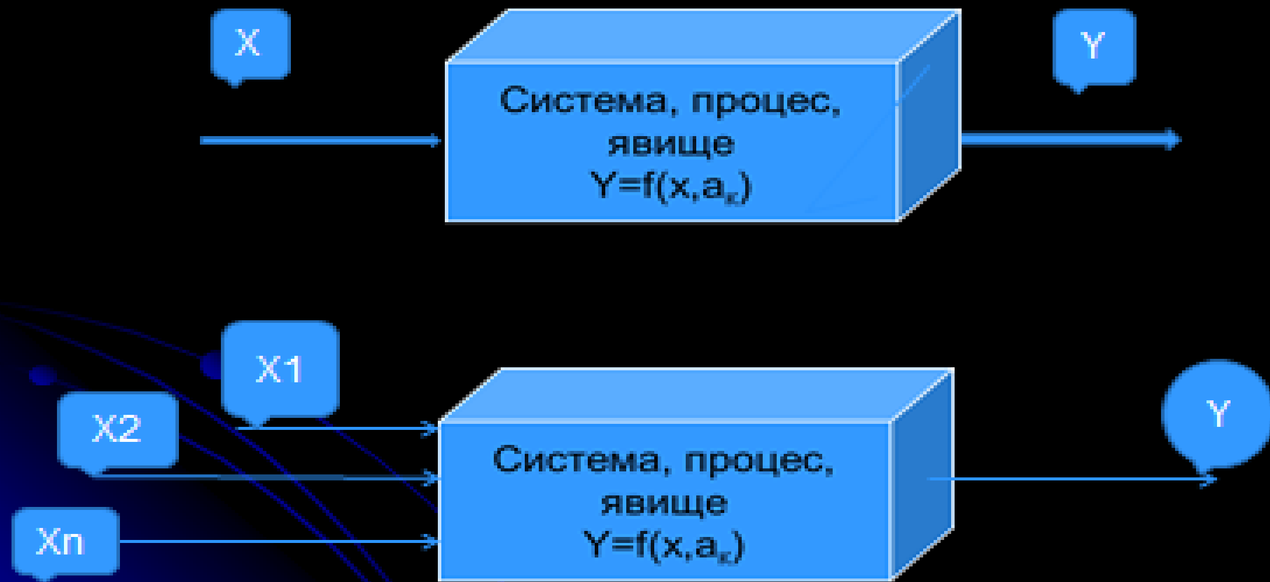
Мета 3:

- *виявлення причинних зв'язків між змінними X та Y.*

Така задача передбачає проникнення у «фізичний механізм» вивчаємих статистичних зв'язків, тобто у механізм перетворення вхідних впливів X та E в результуючі показники Y. Тому, головною метою – є структура функції $f(x)$, яка часто залежить від параметрів, що мають визначену «фізичну» інтерпретацію.

Задачі апроксимації.

- Функції виду $\overline{y} = f(\overline{x}, \overline{a})$ виражають співвідношення “вхід–вихід” будь-якої системи ($\overline{x}$ – вектор вхідних факторів; $\overline{a}$ – вектор параметрів моделі, що належить <u>визначенню</u> за експериментальними даними; $\overline{y}$ – вектор відгуків – вихідні фактори).



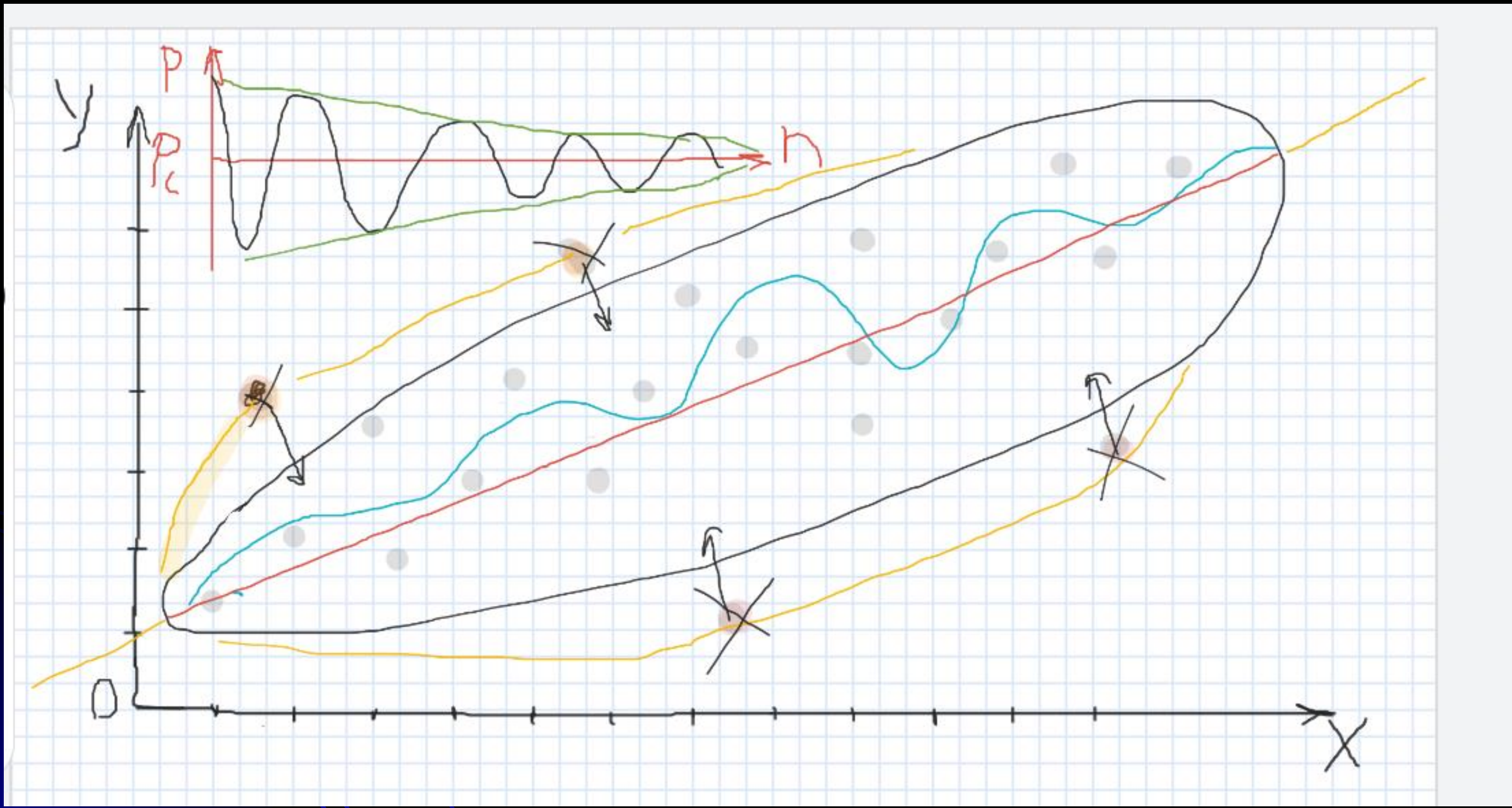
Задачі апроксимації

Задачі апроксимації

- Обробка експериментальних даних (апроксимація даних) в залежності від мети передбачає вирішування наступних задач:
- задача інтерполяції – побудова безперервної функції , що з'єднує всі експериментальні точки ;
- задача екстраполяції – побудова функції за границями відомого інтервалу значень – прогноз;
- задача регресії – побудова найближеної (усередненої) функції
- задача фільтрації (згладжування) – побудова апроксимуючої функції для зниження систематичної похибки експериментальних даних.

- При інтерполяції доцільно з'єднувати експериментальні точки не ламаною лінією, а згладженою кривою. Тому частіше використовують сплайн – інтерполяцію (лінійну, квадратичну, кубічну)

- Всі функції згладжування містять аргументом масиви даних і видають в результаті вектор згладжених даних. Тому, доцільно виконувати згладжування разом з інтерполяцією або регресією.


P
n
Y
X
0

3. Методи та алгоритми: метод найменших квадратів, «ковзне середнє», експоненціальне згладжування

- У дослідженні процесів і систем набули широкого використання математичні моделі, які містять різні функціональні залежності

- Щоб математичні моделі адекватно описували процеси і системи, необхідно використовувати досить адекватні функціональні залежності (математичні формули).
- Таким чином, важливого значення набувають методи апроксимації – методи наближеного зображення реальних функцій такими стандартними аналітичними виразами, як, наприклад, алгебраїчні багаточлени.

$$P(x) = \sum_{K=0}^{m-1} a_K x^K, (m = 1, 2, \ldots).$$

В розгорнутому вигляді можна записати:

$$P(x) = a_0 x^0 + a_1 x^1 + a_2 x^2 + \ldots + a_{m-1} x^{m-1}$$

Важливе достоїнство апроксимуючих багаточленів *P(x)* – їх лінійність відносно невідомих коефіцієнтів a_k (які треба знайти для апроксимації), що дозволяє будувати ефективні алгоритми наближення за допомогою таких функцій.

Один із методів математичної обробки даних досліду (спостережень) - є метод найменших квадратів (МНК), результатом застосування якого є отримання числових коефіцієнтів емпіричної формули (вибраного багаточлена *P(x)*).

Отже, побудова оптимального апроксимуючого багаточлена *P(x)* зводиться до знаходження коефіцієнтів, які мінімізують функцію

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

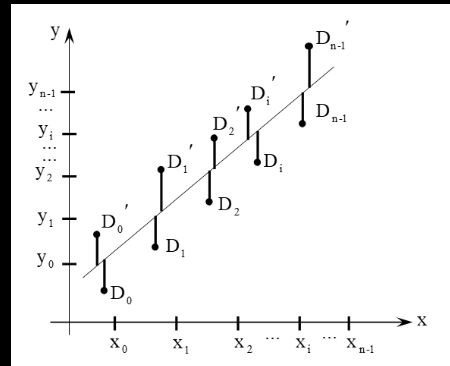
Алгоритм

1. Обирають апроксимуючу функцію

2. Знаходять систему частинних похідних і прирівнюють їх до нуля

3. Розв’язують систему рівнянь відносно коефіцієнтів a_k

Ю. Бродський



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. 2.15).

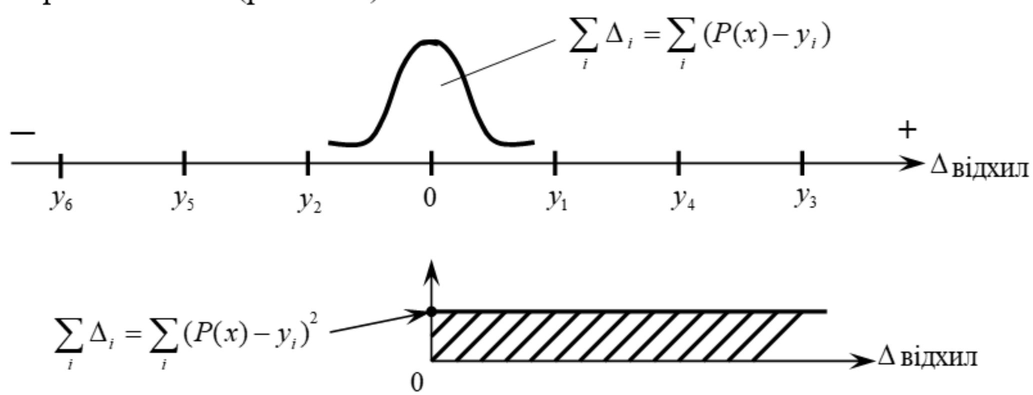


Рис. 2.15

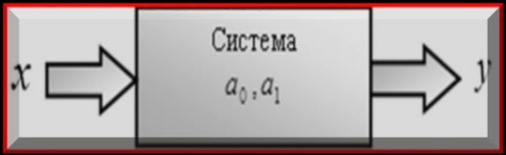
Уникнути цього недоліку дозволяє міра відхилення (критерій), яка була запропонована французьким математиком Лежандром (і паралельно Гауссом) у 1806 році, – брати суму квадратів відхилень.

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

Розглянемо задачу обчислення коефіцієнтів для лінійної функції $P(x_i)=a_0+a_1x_i$

Підставляємо у формулу:

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



Розкриваємо дужки

$$\left[\left(a_0+a_1x_i\right)-y_i\right]^2=\left(a_0+a_1x_i\right)^2-2\left(a_0+a_1x_i\right)y_i+y_i^2=$$
$$=a_0^2+2a_0a_1x_i+a_1^2x_i^2-2a_0y_i-2a_ix_iy_i+y_i^2.$$

Запишемо систему частинних похідних і прирівняємо їх до нуля:

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

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

$$\begin{cases} na_0+a_1\sum\limits_{i=0}^{n-1}x_i=\sum\limits_{i=0}^{n-1}y_i \\ a_0\sum\limits_{i=0}^{n-1}x_i+a_1\sum\limits_{i=0}^{n-1}x_i^2=\sum\limits_{i=0}^{n-1}x_iy_i \end{cases}$$

Розв’язуючи систему відносно a_0 та a_1, маємо:

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

Методи лінійного згладжування («ковзне середнє»)

$$y_{\phi_i} = \frac{y_i + y_{i+1} + \ldots + y_{i+k-1}}{\Delta t},$$

та експоненціального згладжування

$$\hat{y}_{\phi_{i+1}} = \alpha y_i + (1-\alpha)\hat{y}_{\phi_i}$$

4. Інструментарій апроксимації даних в Excel та Mathcad

4.1. Інструментарій програми Excel

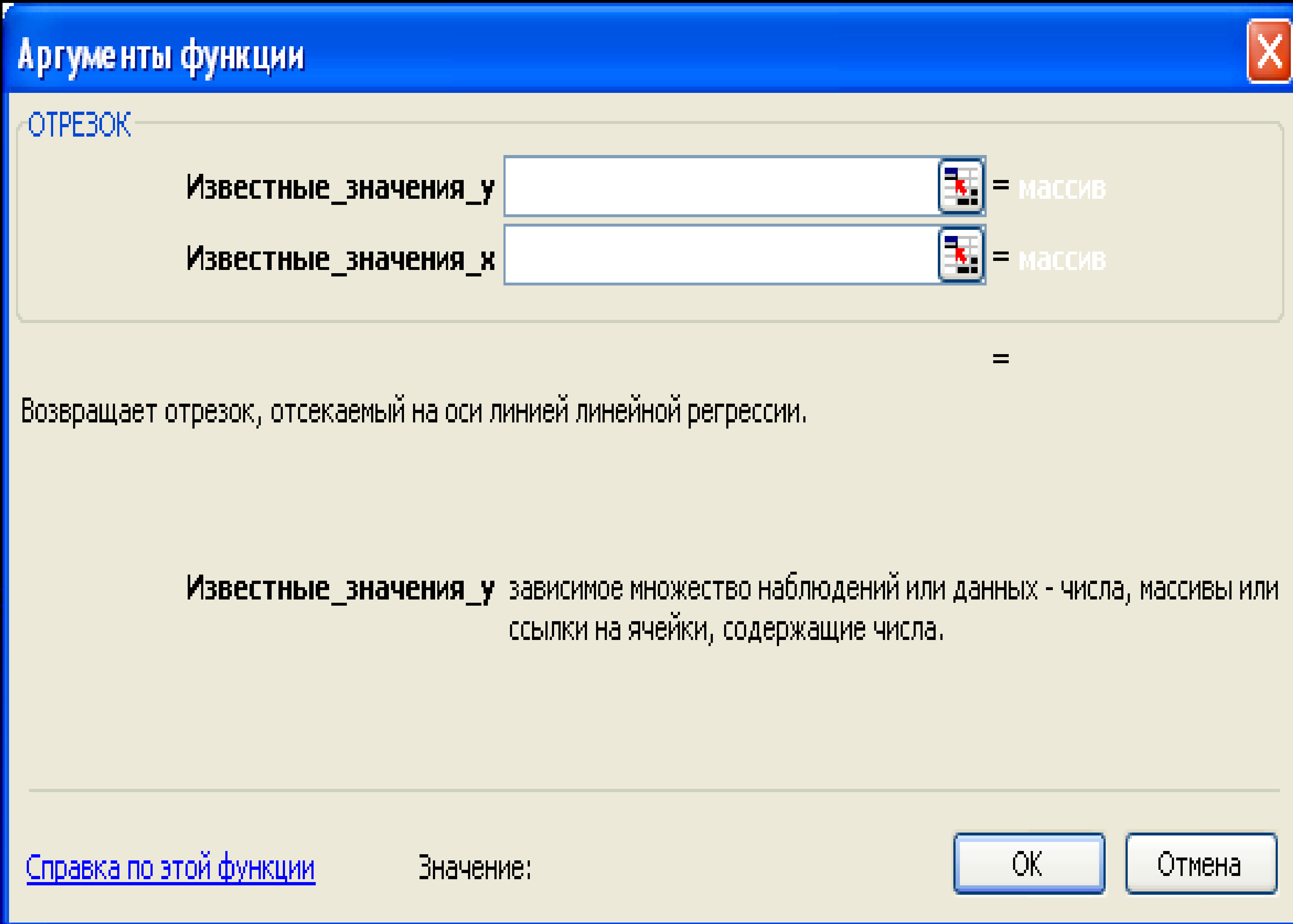
4.1.1. Задачі інтерполяції, регресії, фільтрації (згладжування):

Алгоритм МНК

- Функції категорії “статистичні”
- Побудова “лінії тренду” (графічний спосіб)
- “Пакет аналізу”

Інші алгоритми

- ковзне середнє
- експоненціальне згладжування


Аргументы функции
ОТРЕЗОК
Известные_значения_y
= массив
Известные_значения_x
= массив
=
Возвращает отрезок, отсекаемый на оси линией линейной регрессии.
Известные_значения_y зависимое множество наблюдений или данных - числа, массивы или ссылки на ячейки, содержащие числа.
Справка по этой функции
Значение:
ОК
Отмена

Аргументы функции

НАКЛОН

Известные_значения_y = массив

Известные_значения_x = массив

=

Возвращает наклон линии линейной регрессии.

Известные_значения_y массив или диапазон, содержащий числовые зависимые элементы данных.

Справка по этой функции

Значение:

ОК Отмена

Аргументы функции

КВПИРСОН

Известные_значения_y = массив

Известные_значения_x = массив

=

Возвращает квадрат коэффициента корреляции Пирсона по данным точкам.

Известные_значения_y массив или диапазон, могущий включать числа или имена, массивы или ссылки на ячейки с числами.

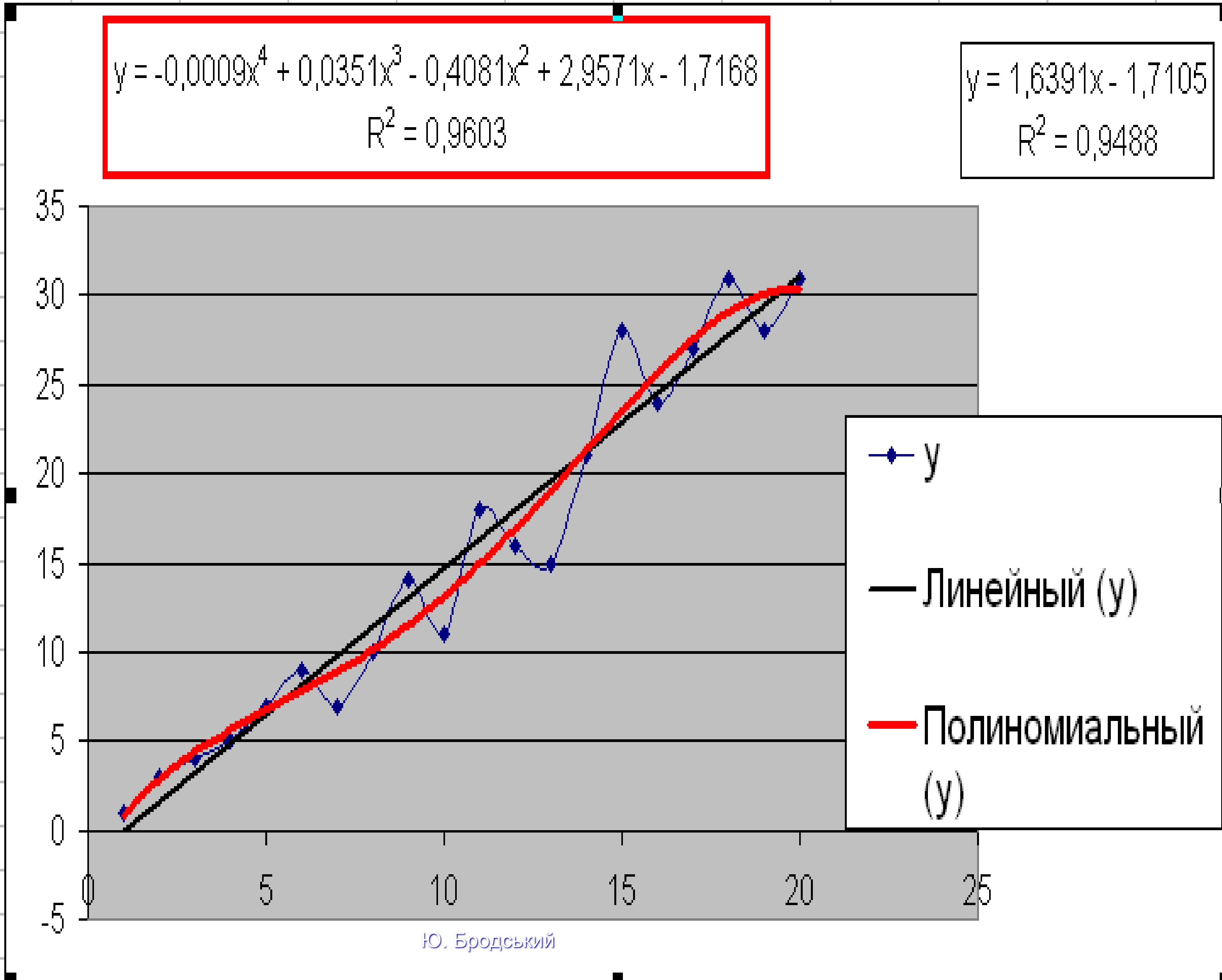
Справка по этой функции

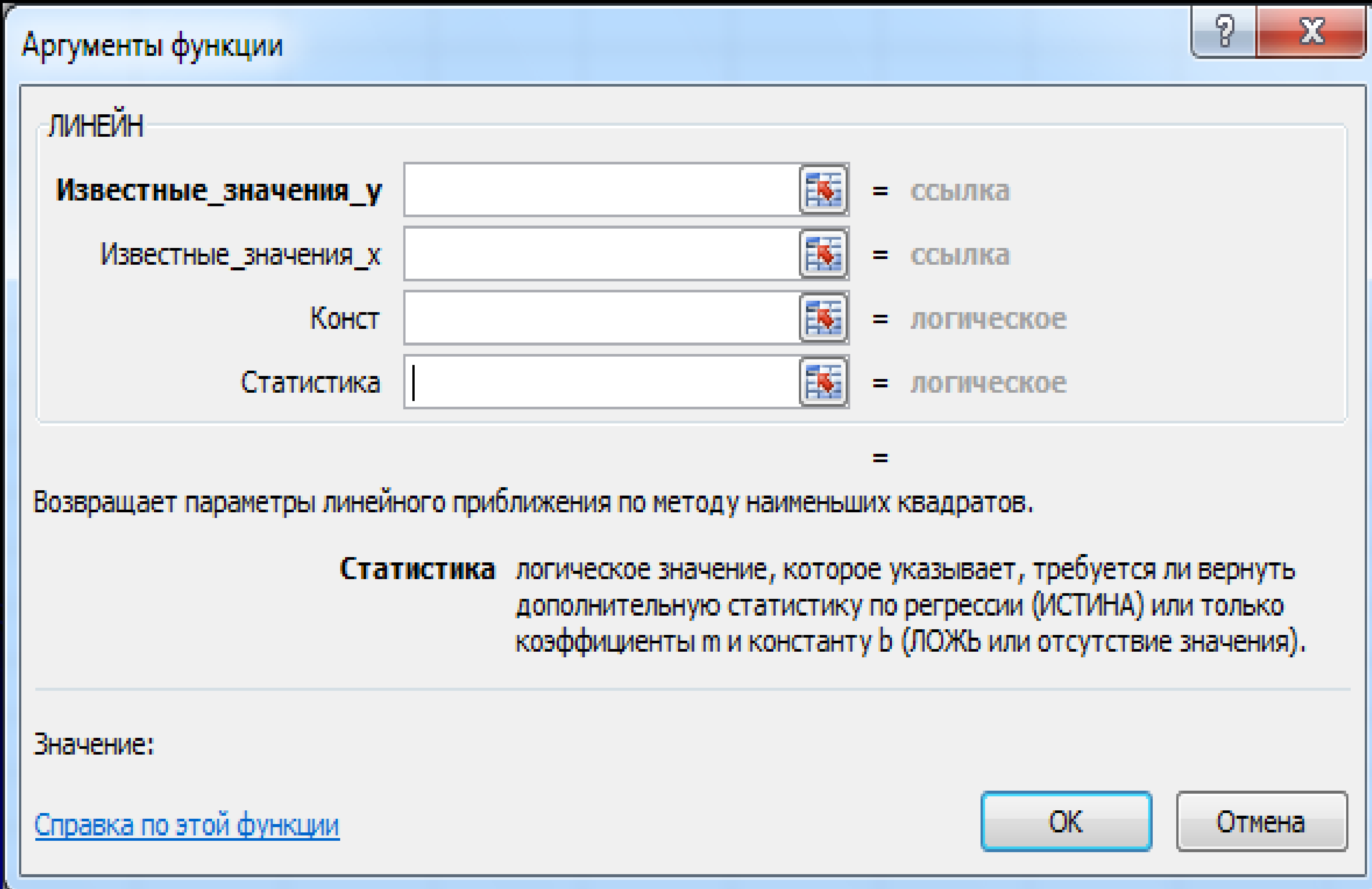
Значение:

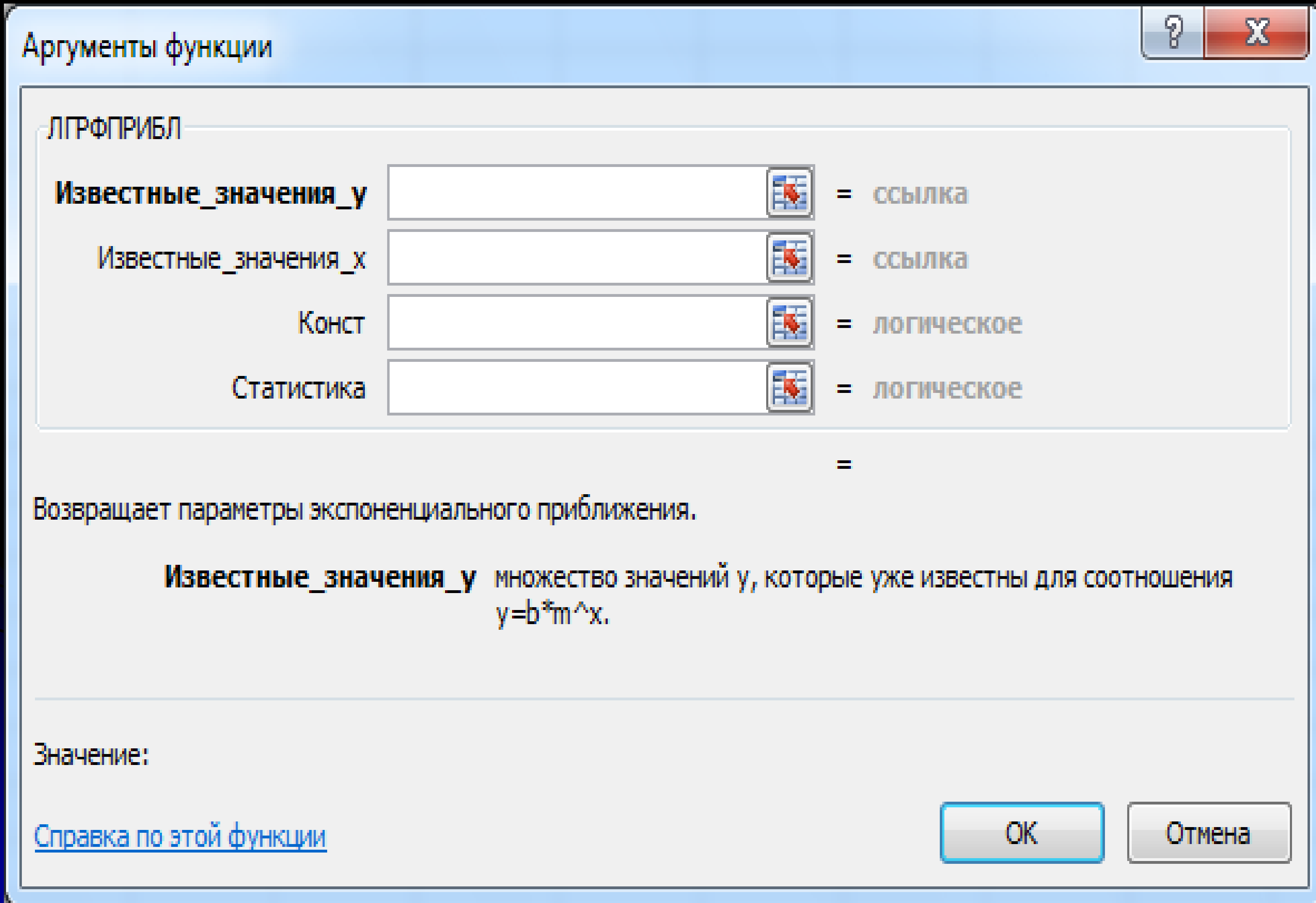
OK

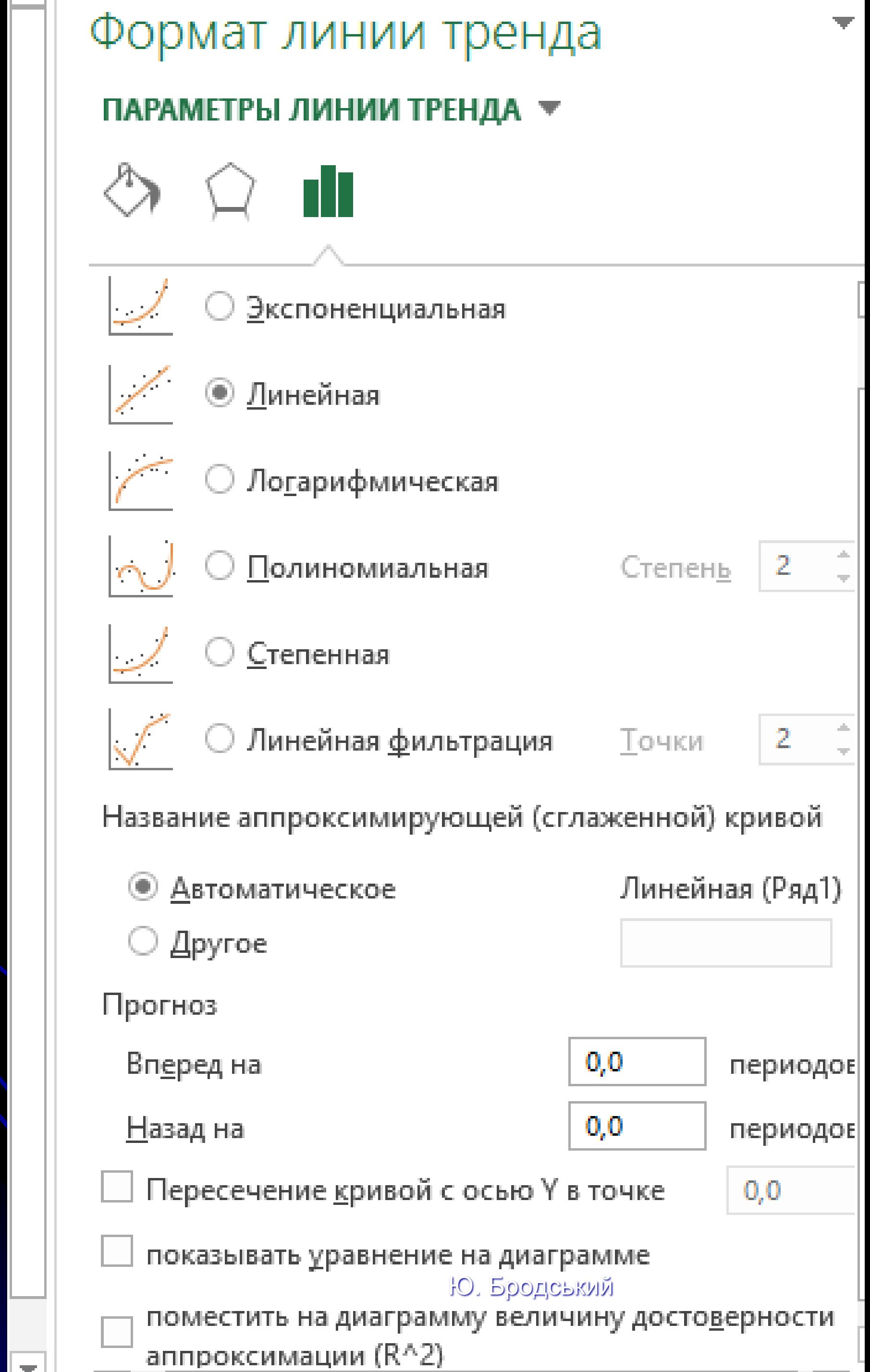
Отмена

x	y
1	1
2	3
3	4
4	5
5	7
6	9
7	7
8	10
9	14
10	11
11	18
12	16
13	15
14	21
15	28
16	24
17	27
18	31
19	28
20	31



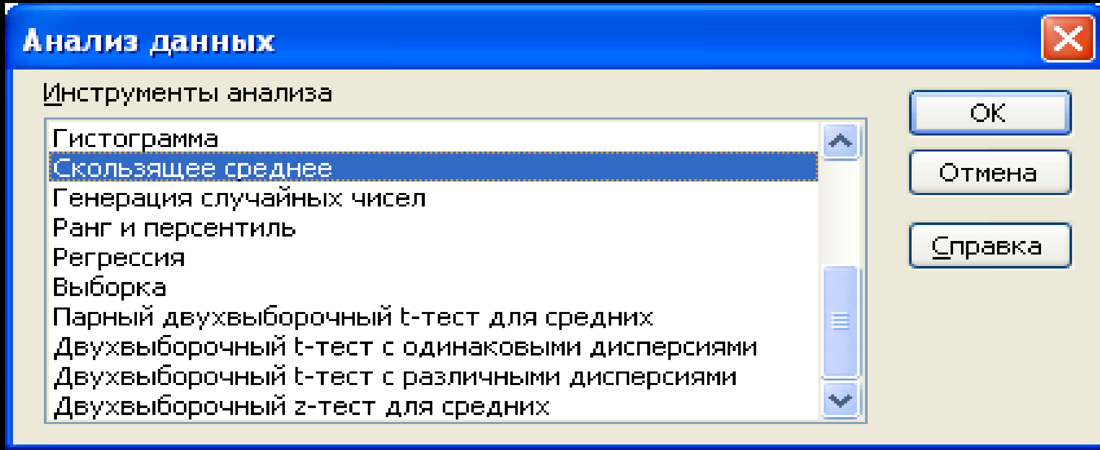

Аргументы функции
ЛИНЕЙН
Известные_значения_y
= ссылка
Известные_значения_x
= ссылка
Конст
= логическое
Статистика
= логическое
=
Возвращает параметры линейного приближения по методу наименьших квадратов.
Статистика
логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии (ИСТИНА) или только коэффициенты m и константу b (ЛОЖЬ или отсутствие значения).
Значение:
Справка по этой функции
ОК
Отмена

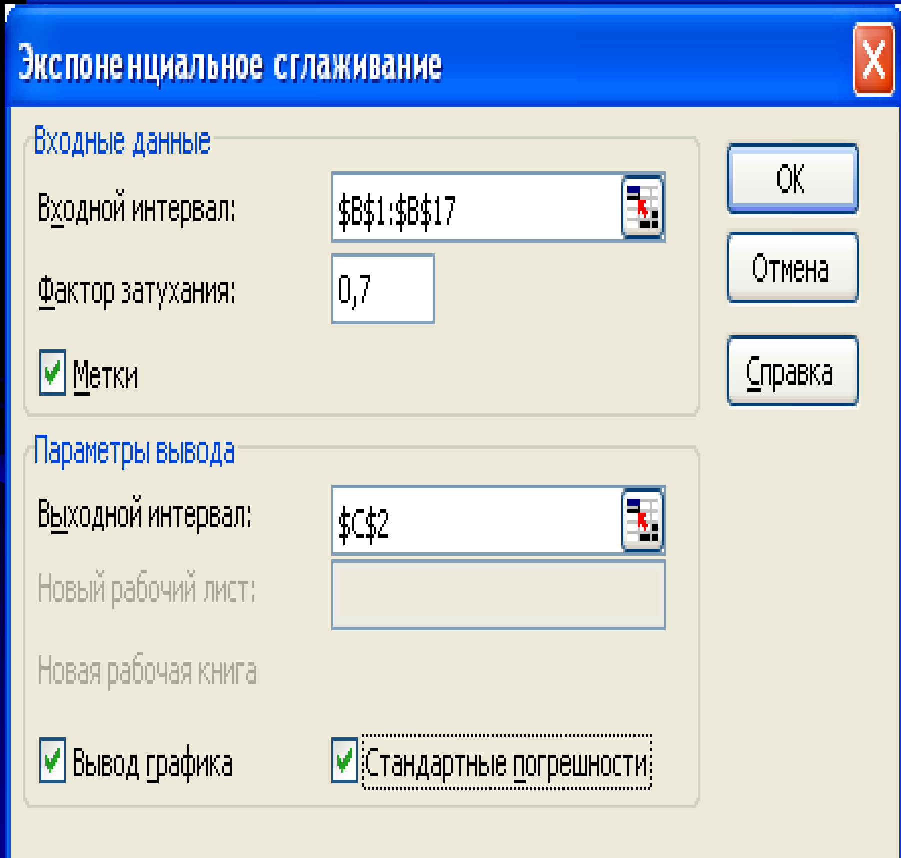

Аргументы функции
ЛГРФПРИБЛ
Известные_значения_y
= ссылка
Известные_значения_x
= ссылка
Конст
= логическое
Статистика
= логическое
=
Возвращает параметры экспоненциального приближения.
Известные_значения_y множество значений y, которые уже известны для соотношения y=b*m^x.
Значение:
Справка по этой функции
ОК
Отмена

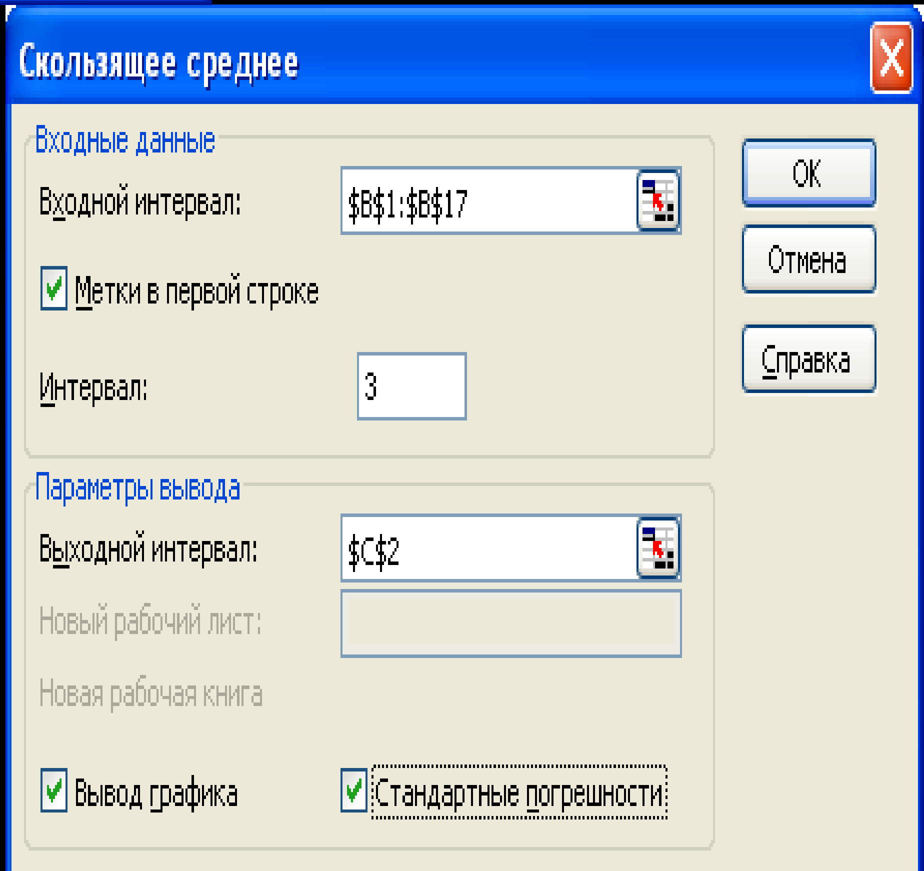

Формат линии тренда
ПАРАМЕТРЫ ЛИНИИ ТРЕНДА
Экспоненциальная
Линейная
Логарифмическая
Полиномиальная
Степень
2
Степенная
Линейная фильтрация
Точки
2
Название аппроксимирующей (сглаженной) кривой
Автоматическое
Линейная (Ряд1)
Другое
Прогноз
Вперед на
0,0
периодов
Назад на
0,0
периодов
Пересечение кривой с осью Y в точке
0,0
показывать уравнение на диаграмме
Ю. Бродський
поместить на диаграмму величину достоверности аппроксимации (R^2)

Ю. Бродський

Пакет аналізу

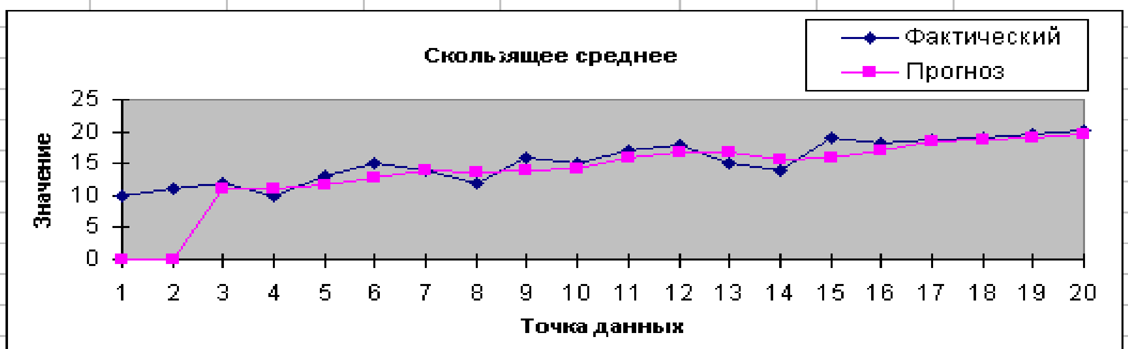


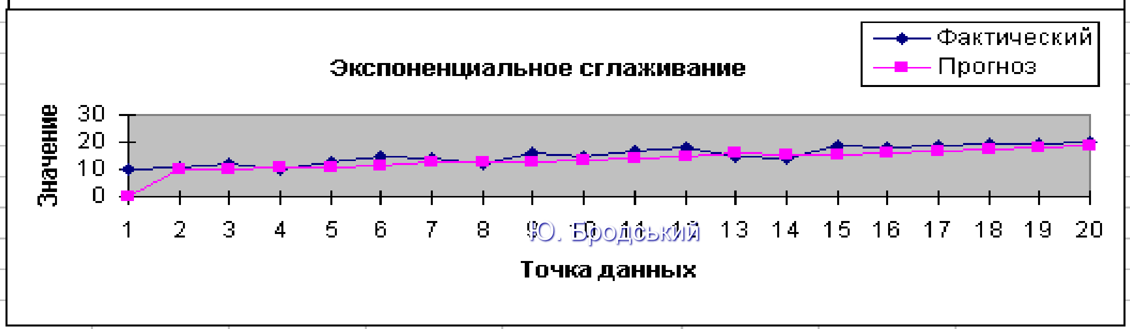


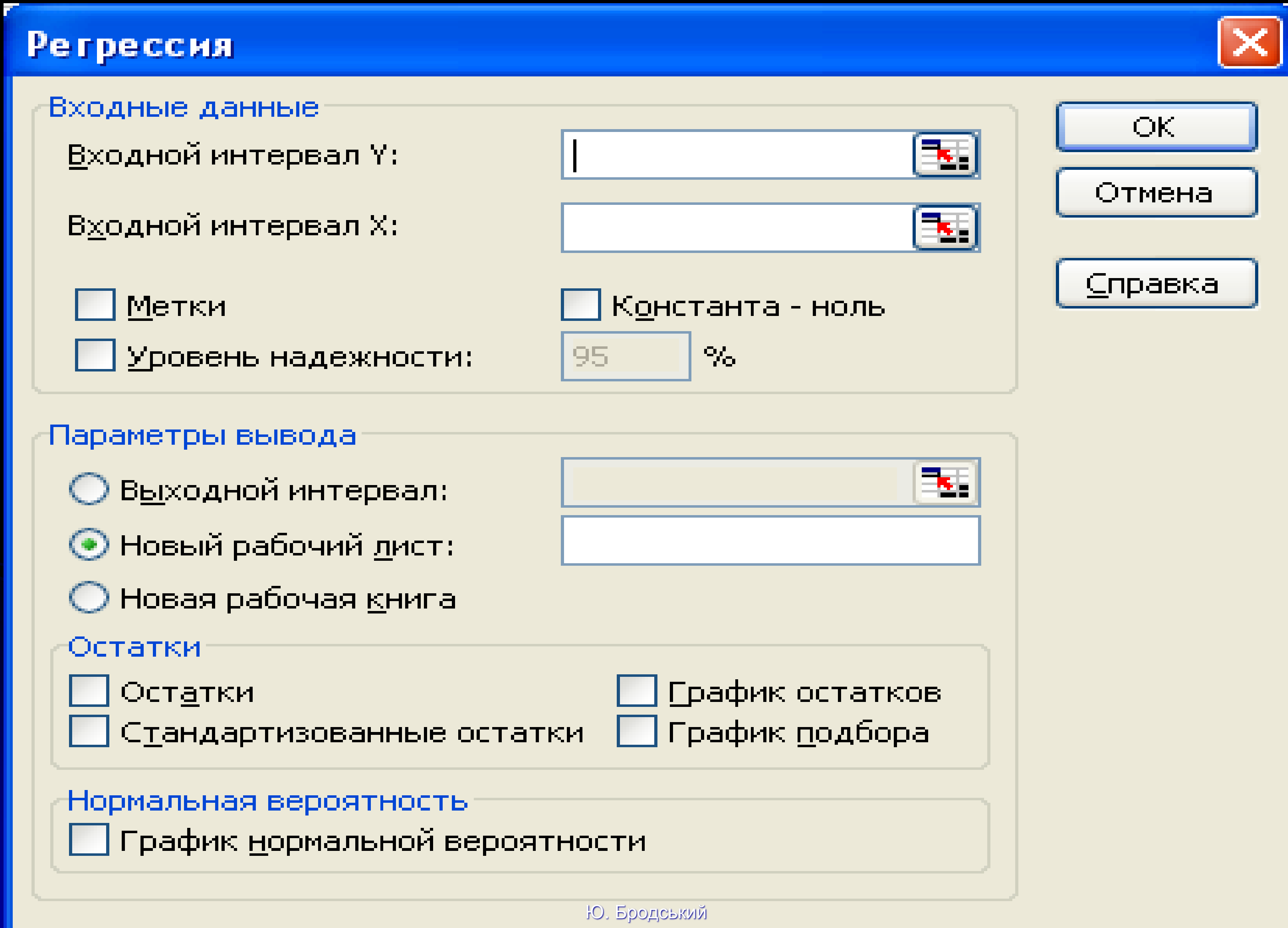


G4 f_x =0,3*B3+0,7*G3

	A	B	C	D	E	F	G	H
1	x	y		$y_{ковз.сер.}$	Стандартна похибка		$y_{експ.згл.}$	Стандартна похибка
2	1	10		#Н/Д	#Н/Д		#Н/Д	#Н/Д
3	2	11		#Н/Д	#Н/Д		10	#Н/Д
4	3	12		11	#Н/Д		10,3	#Н/Д
5	4	10		11	#Н/Д		10,81	#Н/Д
6	5	13		11,66667	1,122167215		10,567	1,231002302
7	6	15		12,66667	1,65551827		11,2969	1,776287233
8	7	14		14	1,551582227		12,40783	2,600547289
9	8	12		13,66667	1,65551827		12,88548	2,718298728
10	9	16		14	1,503083251		12,61984	2,382717742
11	10	15		14,33333	1,551582227		13,63389	2,216948792
12	11	17		16	1,347150628		14,04372	2,16609086
13	12	18		16,66667	1,03637545		14,9306	2,709942421
14	13	15		16,66667	1,360827635		15,85142	2,583734489
15	14	14		15,66667	1,56347192		15,596	2,509030635
16	15	19		16	2,202691956		15,1172	2,056965057
17	16	18,12381		17,04127	2,077632462		16,28204	2,473073775
18	17	18,63095		18,58492	1,841558315		16,83457	2,646726481
19	18	19,1381		18,63095	0,690701326		17,37348	2,689192672
20	19	19,64524		19,1381	0,414932395		17,90287	1,801200105
21	20	20,15238		19,64524	0,507142857		18,42558	1,767927042






Регрессия
Входные данные
Входной интервал Y:
Входной интервал X:
Метки
Константа - ноль
Уровень надежности:
95
%
Параметры вывода
Выходной интервал:
Новый рабочий лист:
Новая рабочая книга
Остатки
Остатки
Стандартизованные остатки
График остатков
График подбора
Нормальная вероятность
График нормальной вероятности
OK
Отмена
Справка
Ю. Бродський

4.1.2. Задача екстраполяції

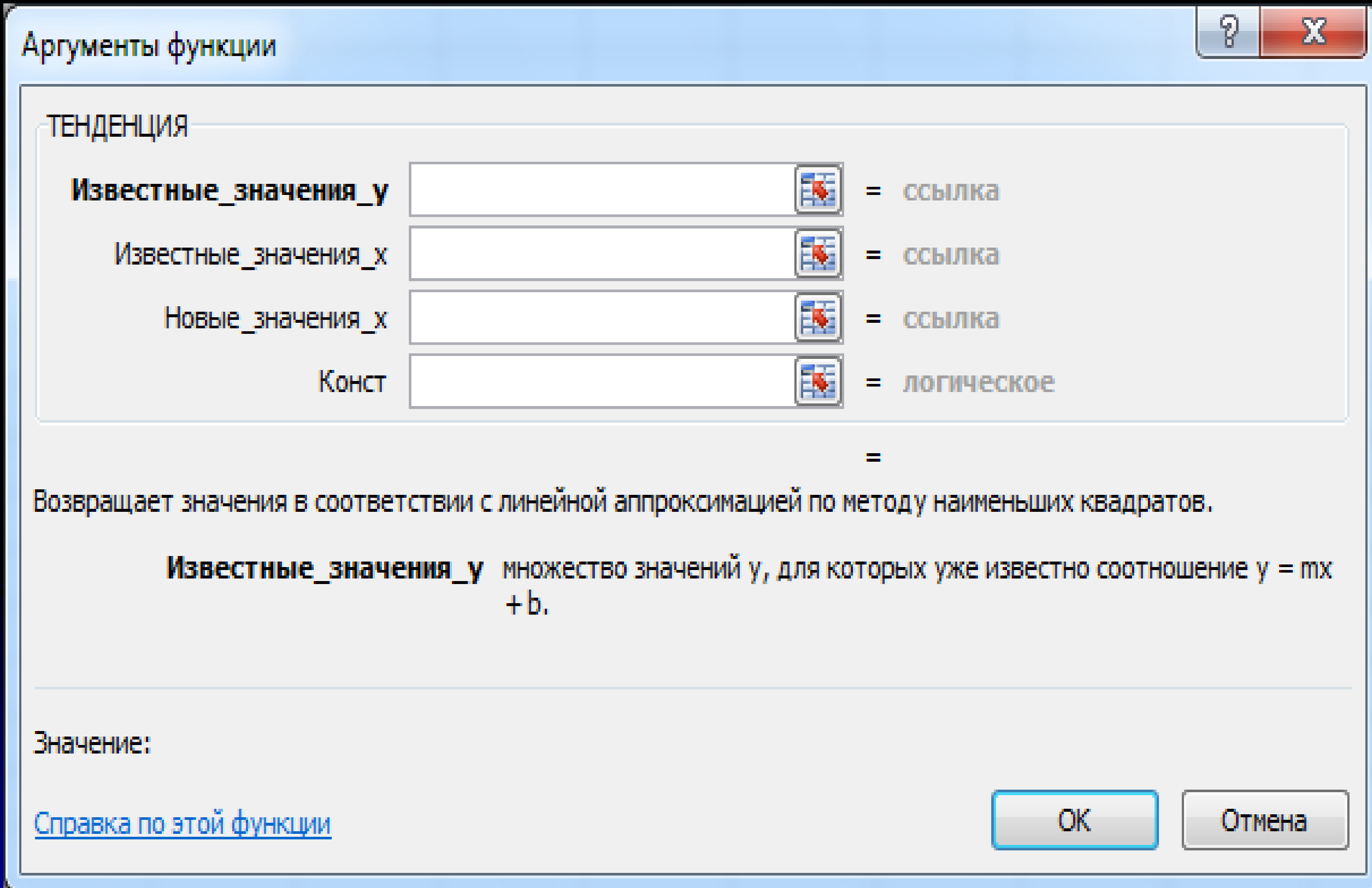
Алгоритм МНК

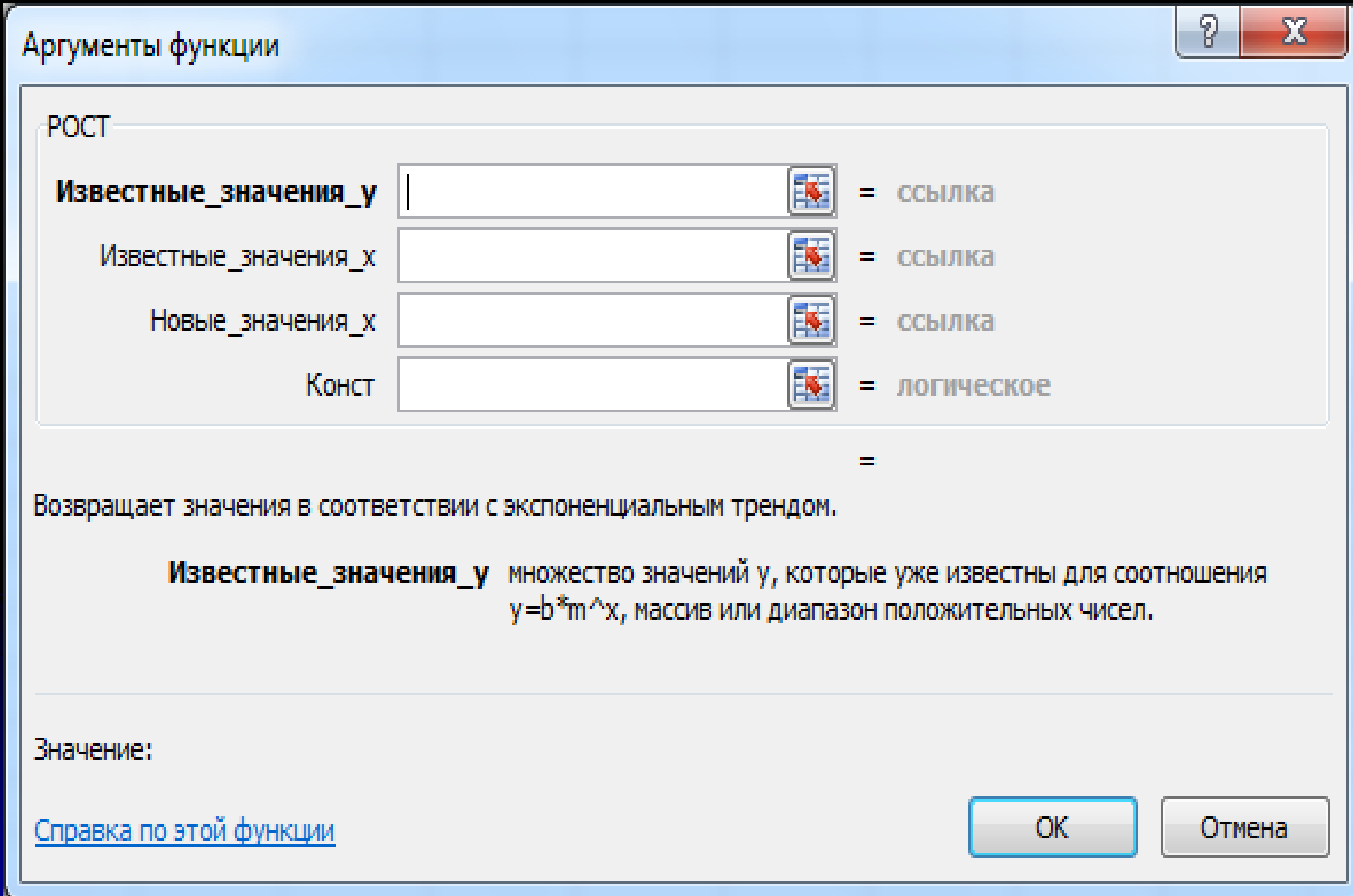
- Функції категорії "статистичні" ТЕНДЕНЦИЯ, ПРЕДСКАЗ, РОСТ
- Побудова "лінії тренду" (графічний спосіб)
- "Пакет аналізу"
- "Маркер заповнення" (*ЛКМ*) – лінійне наближення
- "Маркер заповнення" (*ПКМ*) – лінійне наближення або експоненціальне згладжування

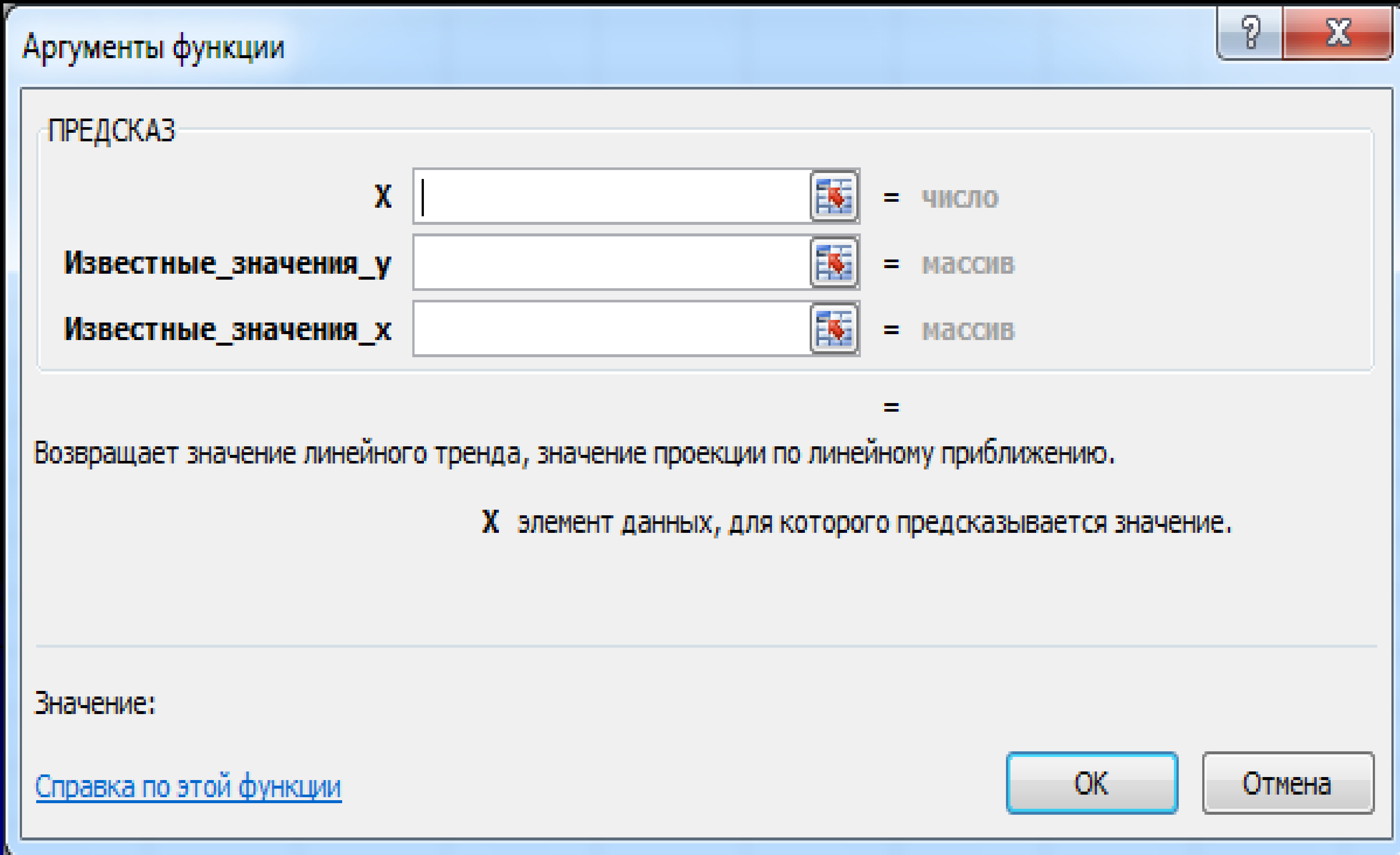
- Функції «**Тенденция**» та «**Предсказ**» апроксимують прямою лінією, а функція «**Рост**» експонентою (за методом найменших квадратів)

$$"Тенденция" \Rightarrow \overline{y} = a_0 + a_1{}^{*}\overline{x};$$
$$"Рост" \Rightarrow \overline{y} = a_0{}^{*}\exp\left(a_1\overline{x}\right).$$

- Параметри функцій однакові (для обох функцій): **ТЕНДЕНЦИЯ** (значення Y; старі значення X ; нові значення X для прогнозу; логічна константа: якщо «1» - a_o обчислюється, якщо «0» - $a_o = 0$).


Аргументы функции
ТЕНДЕНЦИЯ
Известные_значения_y
= ссылка
Известные_значения_x
= ссылка
Новые_значения_x
= ссылка
Конст
= логическое
=
Возвращает значения в соответствии с линейной аппроксимацией по методу наименьших квадратов.
Известные_значения_y множество значений y, для которых уже известно соотношение y = mx + b.
Значение:
Справка по этой функции
ОК
Отмена


Аргументы функции
РОСТ
Известные_значения_y
= ссылка
Известные_значения_x
= ссылка
Новые_значения_x
= ссылка
Конст
= логическое
=
Возвращает значения в соответствии с экспоненциальным трендом.
Известные_значения_y
множество значений y, которые уже известны для соотношения y=b*m^x, массив или диапазон положительных чисел.
Значение:
Справка по этой функции
ОК
Отмена


Аргументы функции
ПРЕДСКАЗ
X = число
Известные_значения_y = массив
Известные_значения_x = массив
=
Возвращает значение линейного тренда, значение проекции по линейному приближению.
X элемент данных, для которого предсказывается значение.
Значение:
Справка по этой функции
ОК
Отмена

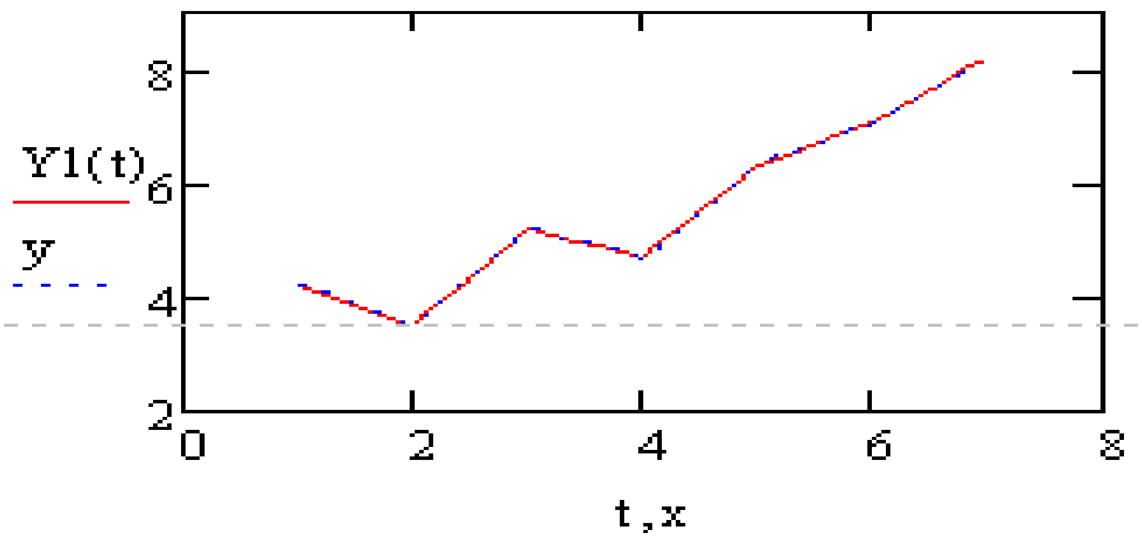
4.2.1. Інтерполяція, екстраполяція, регресія

- **4.2.1.1. Лінійна інтерполяція -** функція *linterp(x,y,t)* , де
- x - вектор значень (даних) аргументу, тобто незалежної змінної;
- y - вектор експериментальних даних (залежної змінної);
- t - значення аргументу, за яким обчислюється інтерполююча функція.

$$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$$

$$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$$

$$Y1(t) := \mathrm{linterp}(x, y, t) \qquad t := 1, 1.2 .. 7$$



4.2.1.2. Сплайн–інтерполяція - перед застосуванням функції *interp(S,x,y,t)* потрібно визначити вектор похідних другого степеня S:

lspline(x,y) - створює вектор значень коефіцієнтів кривої, що наближається до прямої лінії в граничних точках, коефіцієнти лінійного сплайна;

pspline(x,y) - вектор значень коефіцієнтів квадратичного сплайна;

cspline(x,y) - вектор значень кубічного сплайна;

t - значення аргументу, при якому обчислюється функція

$$x := (1 \;\; 2 \;\; 3 \;\; 4 \;\; 5 \;\; 6 \;\; 7)^T$$

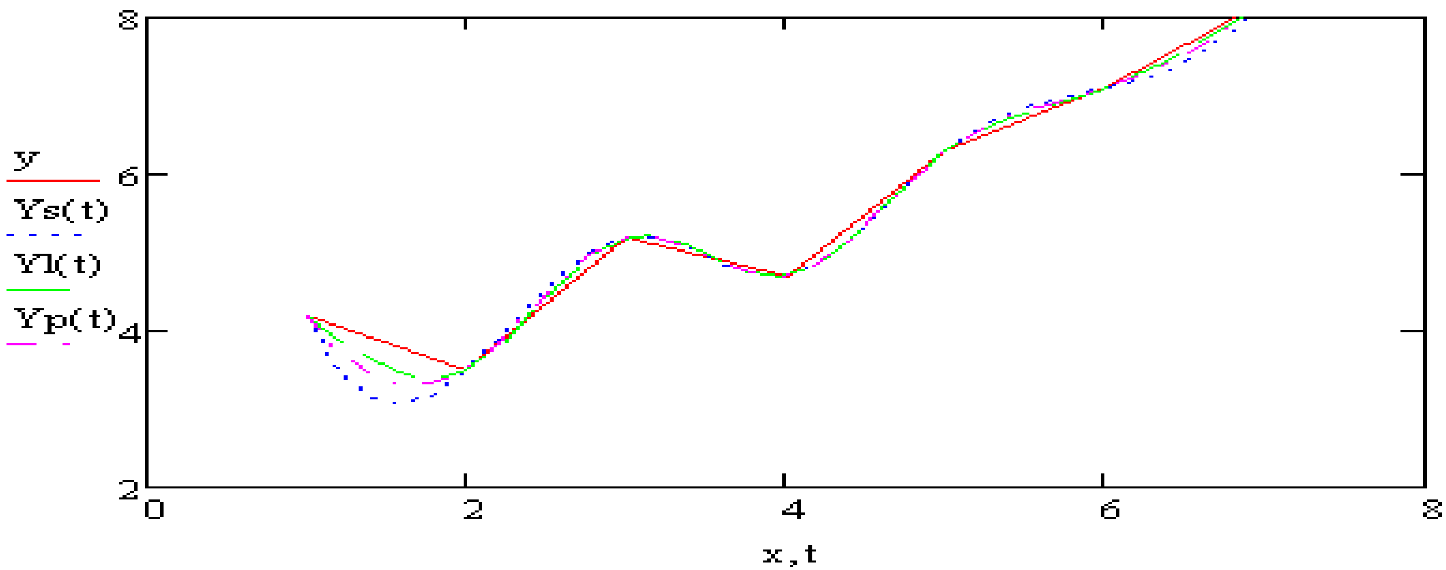
$$y := (4.2 \;\; 3.5 \;\; 5.2 \;\; 4.7 \;\; 6.3 \;\; 7.1 \;\; 8.2)^T$$

$$s := \mathrm{cspline}(x, y)$$

$$Ys(t) := \mathrm{interp}(s, x, y, t)$$

$$Yl(t) := \mathrm{interp}(\mathrm{lspline}(x, y), x, y, t)$$

$$Yp(t) := \mathrm{interp}(\mathrm{pspline}(x, y), x, y, t)$$



4.2.1.3. Сплайн–інтерполяція.

- **Інтерполяція В-сплайнами** – відрізняється способом з’єднання («зшивання») сплайнів не в точках X , а в точках u , координати яких вводить користувач. Вектор значень коефіцієнтів В-сплайна задається функцією: *bspline(x,y,u,n)* ,
- де *u* - вектор значень аргументу, в яких виконується «зшивання» сплайнів – задається користувачем; *n* - порядок поліномів сплайн-інтерполяції (1, 2 або 3).

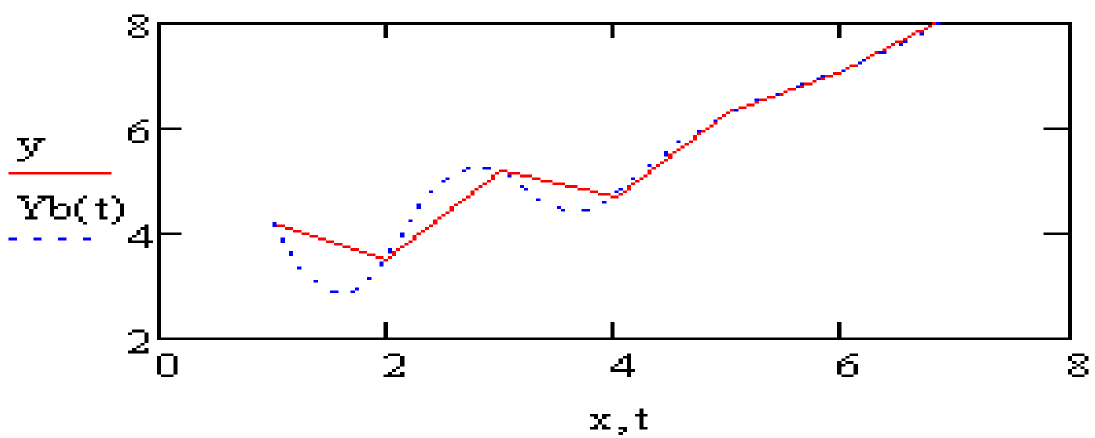
$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

$u := (1\ \ 2.1\ \ 3.2\ \ 4.1\ \ 5.4\ \ 7)^T$

b := bspline(x,y,u,2)

Yb(t) := interp(b,x,y,t)



Екстраполяція даних

функція *predict(y,m,n)*

y - вектор заданих значень функції, аргумент якої змінюється через однакові інтервали;

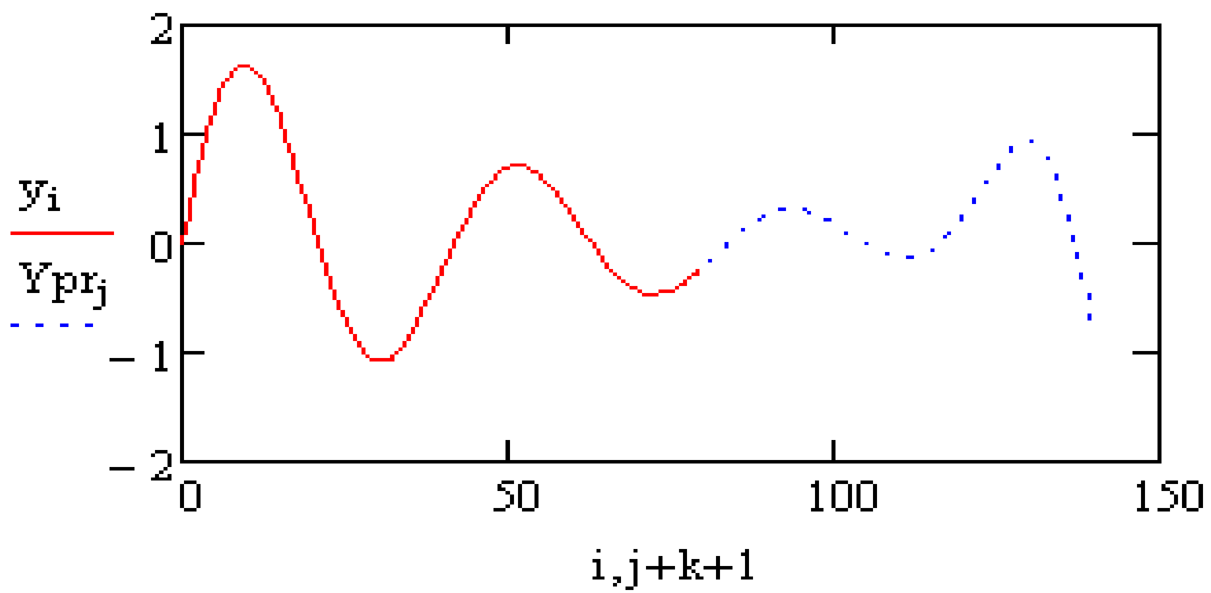
m - кількість послідовних елементів вектора для виконання екстраполяції;

n - кількість точок екстраполяції (прогнозуючих елементів вектора).

$k := 80 \qquad i := 0..k$

$$y_i := 2\cdot e^{-0.02\cdot i}\cdot \sin(0.15\cdot i)$$

$m := 70 \qquad n := 60 \qquad Ypr := predict(y, m, n) \qquad j := 0..n$



Регресія.

- **1.4.1. Лінійна регресія базується на двох методах: МНК і методу медіан.**

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

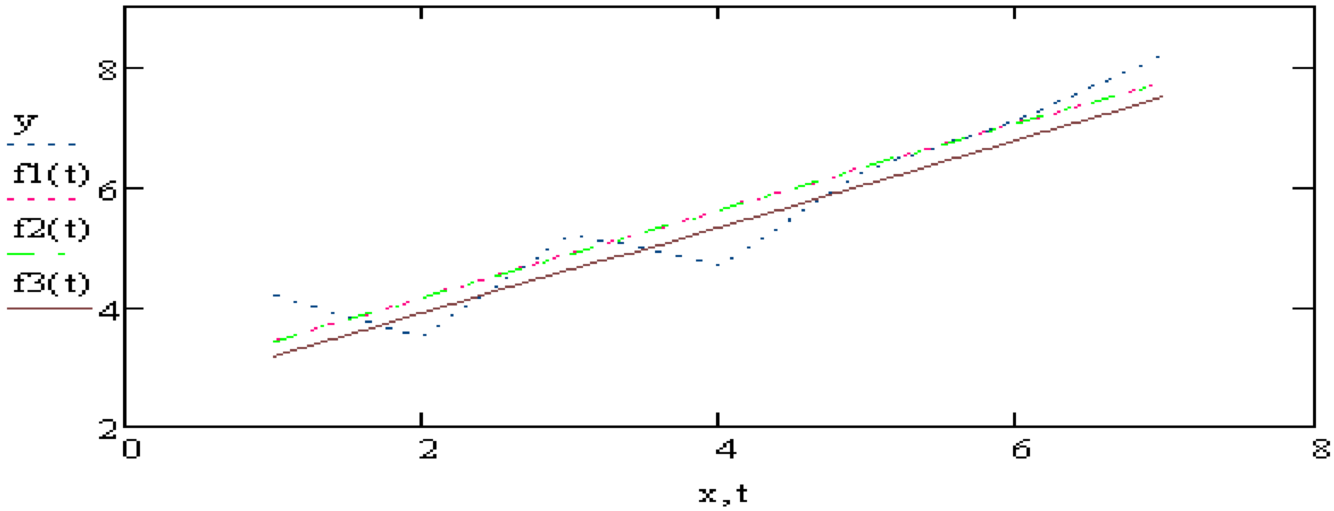
$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

$\mathrm{line}(x,y) = \begin{pmatrix} 2.7 \\ 0.725 \end{pmatrix}$ $\qquad f1(t) := \mathrm{line}(x,y)_0 + \mathrm{line}(x,y)_1 \cdot t$

$\mathrm{intercept}(x,y) = 2.7$ $\qquad f2(t) := \mathrm{intercept}(x,y) + \mathrm{slope}(x,y)\cdot t$

$\mathrm{slope}(x,y) = 0.725$

$\mathrm{medfit}(x,y) = \begin{pmatrix} 2.433 \\ 0.725 \end{pmatrix}$ $\qquad f3(t) := \mathrm{medfit}(x,y)_0 + \mathrm{medfit}(x,y)_1 \cdot t$



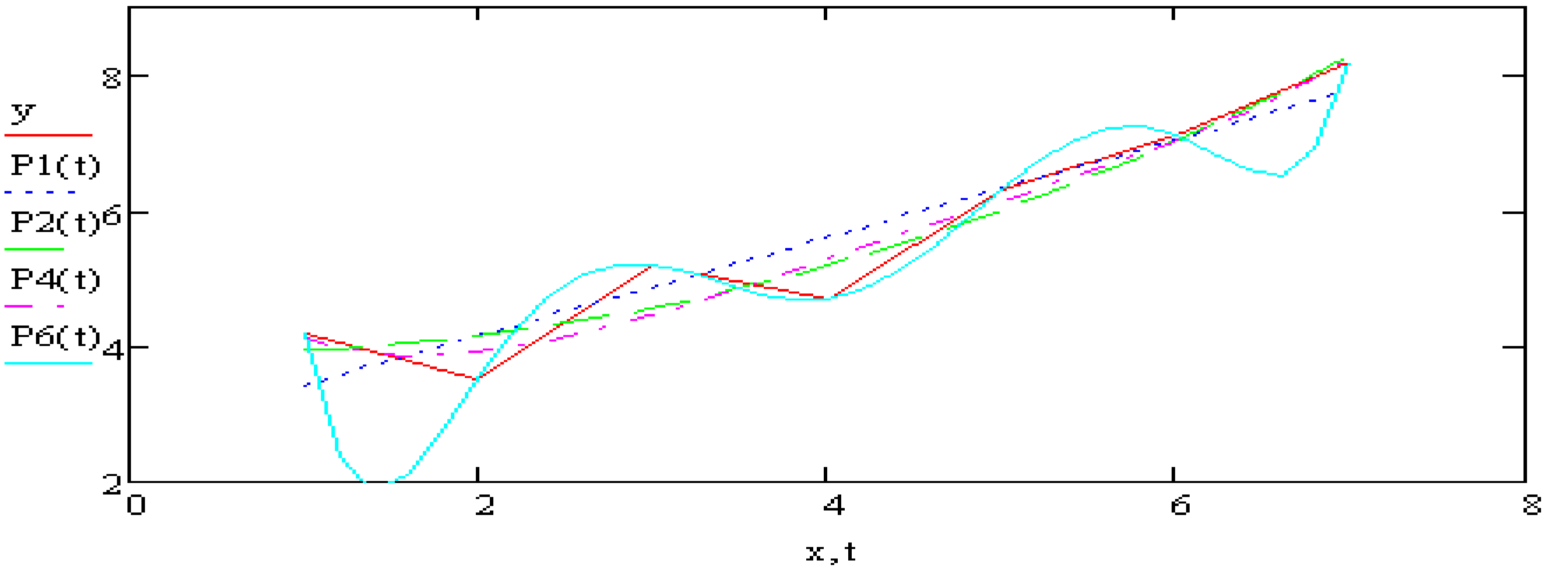
Поліноміальна регресія

- **Спосіб 1. Наближення поліномом k - го степеню ,** $y(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_k x^k$

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

```
k := 1    P1(t) := interp(regress(x,y,k),x,y,t)
k := 2    P2(t) := interp(regress(x,y,k),x,y,t)
k := 4    P4(t) := interp(regress(x,y,k),x,y,t)
k := 6    P6(t) := interp(regress(x,y,k),x,y,t)
```



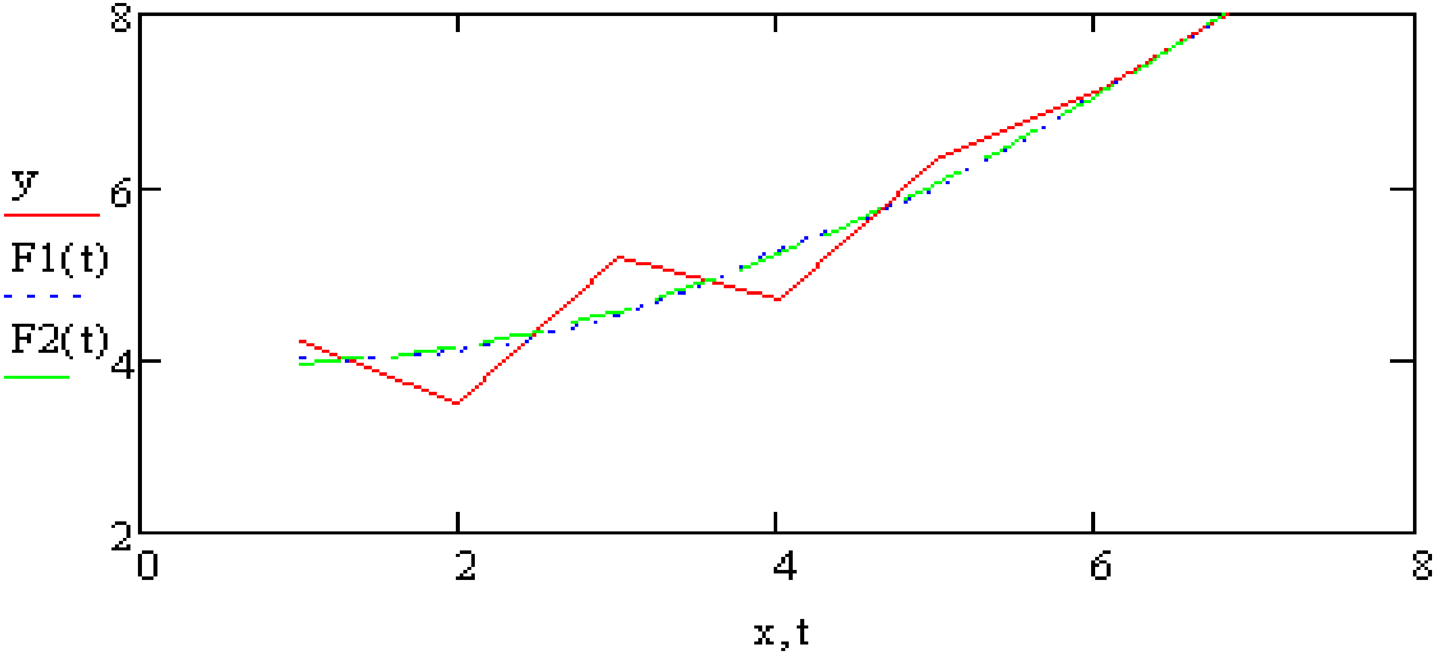
Спосіб 2. Регресія відрізками поліномів

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

$s1 := \mathrm{loess}(x,y,0.9)$ $\qquad$ $s2 := \mathrm{loess}(x,y,2)$

$F1(t) := \mathrm{interp}(s1,x,y,t)$ $\qquad$ $F2(t) := \mathrm{interp}(s2,x,y,t)$



Регресія на базі нелінійних математичних функцій

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

$g := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$C := \mathrm{expfit}(x, y, g)$

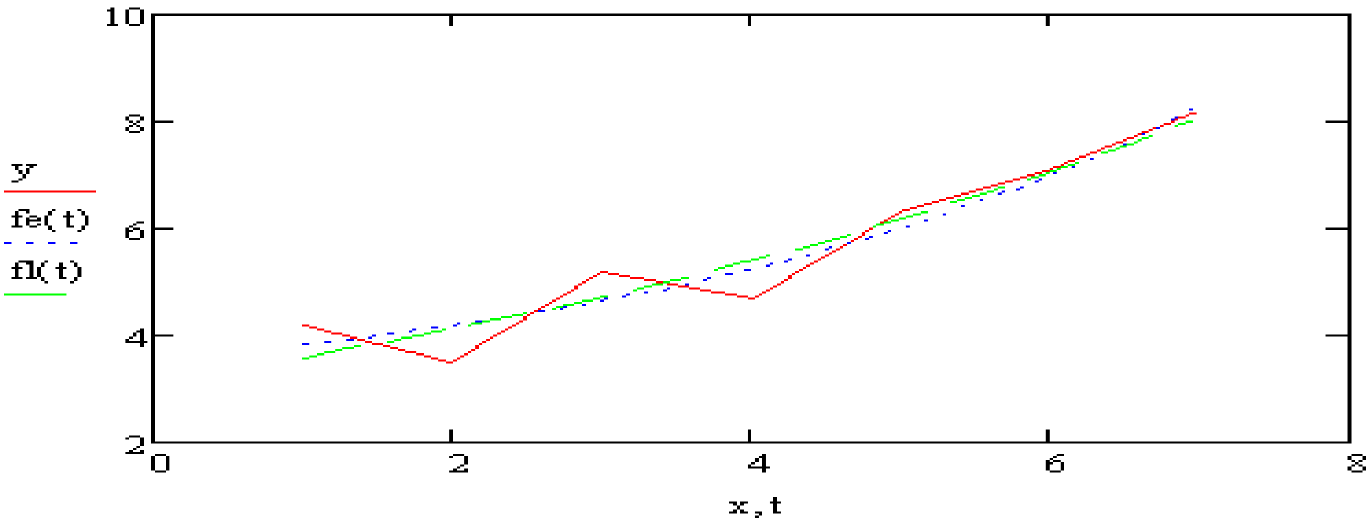
$C = \begin{pmatrix} 0.917 \\ 0.26 \\ 2.646 \end{pmatrix}$

$fe(t) := C_0 \cdot e^{C_1 \cdot t} + C_2$

$D := \mathrm{lgsfit}(x, y, g)$

$D = \begin{pmatrix} 6.893 \times 10^{13} \\ 2.194 \times 10^{13} \\ 0.135 \end{pmatrix}$

$$fl(t) := \frac{D_0}{\left(1 + D_1 \cdot e^{-D_2 \cdot t}\right)}$$



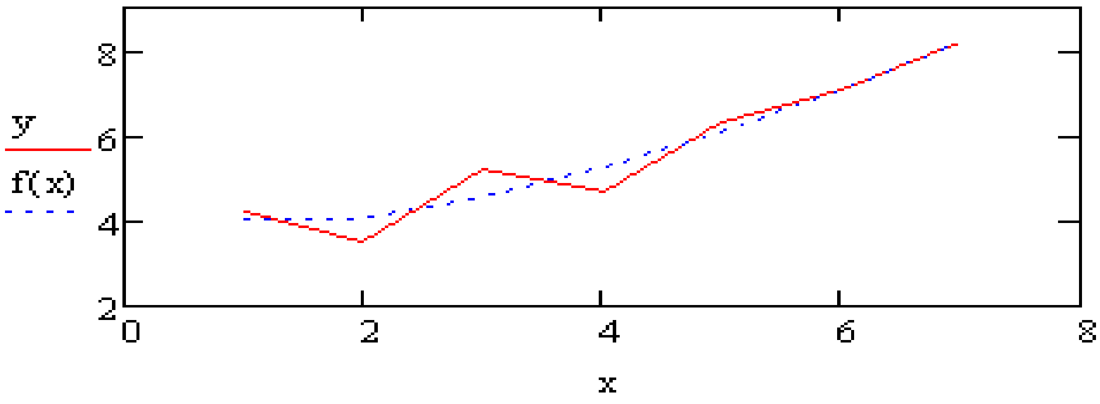
Регресія загального вигляду

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7)^T$

$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2)^T$

$F(x) := \begin{pmatrix} \frac{1}{x+1} \\ e^{0.2x} \\ x \end{pmatrix}$ $\quad C := \mathrm{linfit}(x, y, F) \quad C = \begin{pmatrix} 4.882 \\ 0.768 \\ 0.639 \end{pmatrix}$

$$f(x) := C_1 \cdot e^{0.2 \cdot x} + C_2 \cdot x + \frac{C_0}{x+1}$$



2. Реалізація згладжуючих алгоритмів.

medsmooth(y,b) - згладжування алгоритмом «ковзної медіани», де b - ширина вікна згладжування. Дана функція передбачає, що дані розміщені рівномірно (x - рівномірна послідовність чисел);
ksmooth(x,y,b) - згладжування на основі функції Гауса, де *b* - параметр управління вікном згладжування;
supsmooth(x,y) - згладжування адаптивним алгоритмом за методом найменших квадратів з аналізом найближчих «сусідів» кожної пари даних.

$x := (1\ \ 2\ \ 3\ \ 4\ \ 5\ \ 6\ \ 7\ \ 8\ \ 9\ \ 10\ \ 11\ \ 12\ \ 13\ \ 14\ \ 15)^T$

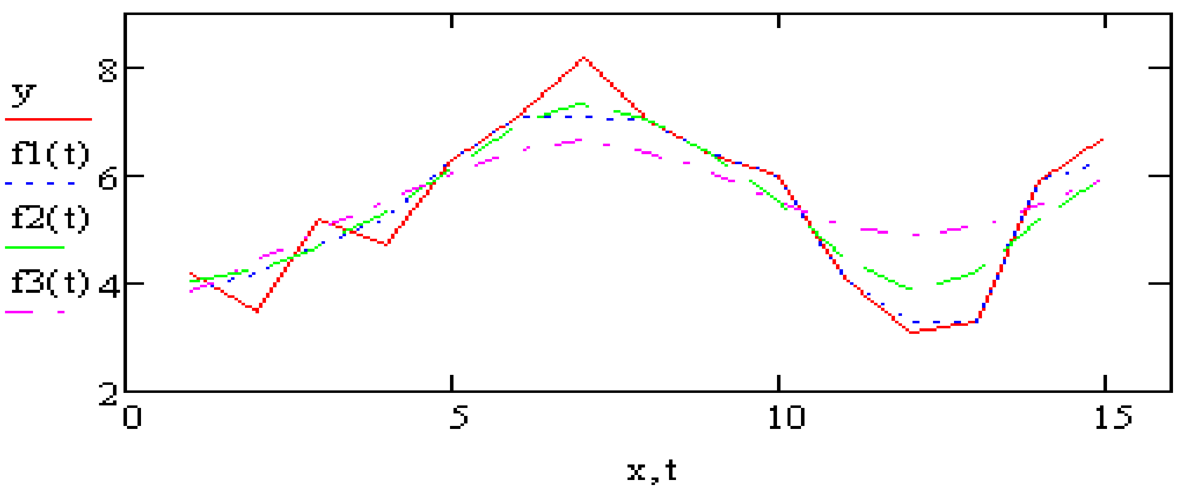
$y := (4.2\ \ 3.5\ \ 5.2\ \ 4.7\ \ 6.3\ \ 7.1\ \ 8.2\ \ 7\ \ 6.4\ \ 6\ \ 4.1\ \ 3.1\ \ 3.3\ \ 5.9\ \ 6.7)^T$

z1 := medsmooth(y,3)   s1 := cspline(x,z1)   f1(t) := interp(s1,x,z1,t)

z2 := ksmooth(x,y,3)   s2 := cspline(x,z2)   f2(t) := interp(s2,x,z2,t)

z3 := supsmooth(x,y)   s3 := cspline(x,z3)   f3(t) := interp(s3,x,z3,t)

t := 1 .. 15



Висновки

- 1). Обробка експериментальних даних (апроксимація даних) в залежності від мети передбачає вирішування наступних задач:
- задача інтерполяції – побудова безперервної функції , що з'єднує всі експериментальні точки ;
- задача екстраполяції – побудова функції за границями відомого інтервалу значень – прогноз;
- задача регресії – побудова найближеної (усередненої) функції
- задача фільтрації (згладжування)– побудова апроксимуючої функції для зниження систематичної похибки експериментальних даних.

- 2). При інтерполяції доцільно з'єднувати експериментальні точки не ламаною лінією, а згладженою кривою. Тому частіше використовують сплайн –інтерполяцію (лінійну, квадратичну, кубічну)

- 3). Всі функції згладжування містять аргументом масиви даних і видають в результаті вектор згладжених даних. Тому, доцільно виконувати згладжування разом з інтерполяцією або регресією.

- 4). Система MathCad на відміну від програми Excel містить набагато більше алгоритмів та, відповідний інструментарій обробки експериментальних даних.

Ю. Бродський