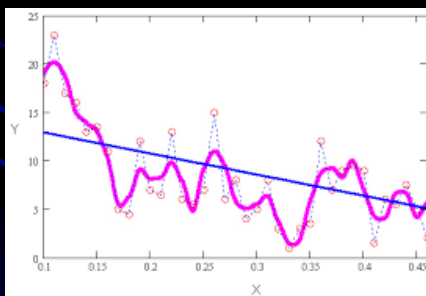




Лекція 5-6 (4 год.)

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.



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

«...Вся проблема цього світу в тому, що дураки та фанатики завжди переконані в собі, а розумні люди повні сумнівів.»

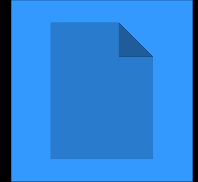
Б. Рассел



Апроксимація (лат. *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
 $Y=f(x, a_k)$

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:

$X = (x_1, x_2, \dots, x_n)^T$, $Y = (y_1, y_2, \dots, y_m)^T$, $E = (e_1, e_2, \dots, e_k)^T$. 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 аналізованих 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 ґрунтується 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 X_i має бути обґрунтований теоретично

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

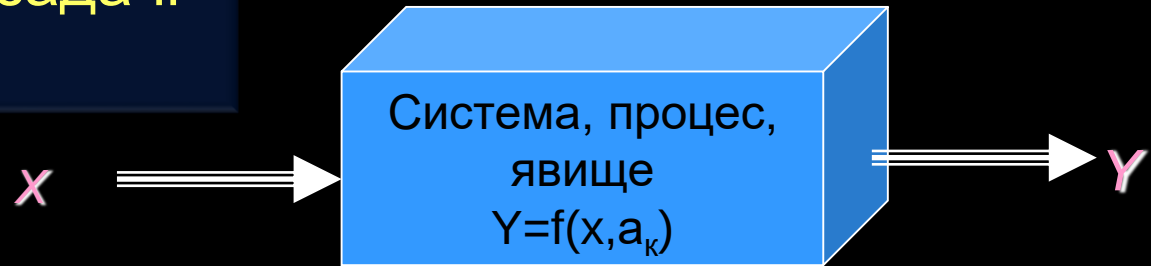
кількість факторів не повинна перевищувати однієї третини числа спостережень у вибірці (довжини часового ряду)

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

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 багатфакторна $y = f(x_1, x_2, \dots, x_n)$

2. Прикладні цілі та задачі апроксимації



- Загальну задачу можна сформулювати так:

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.

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).

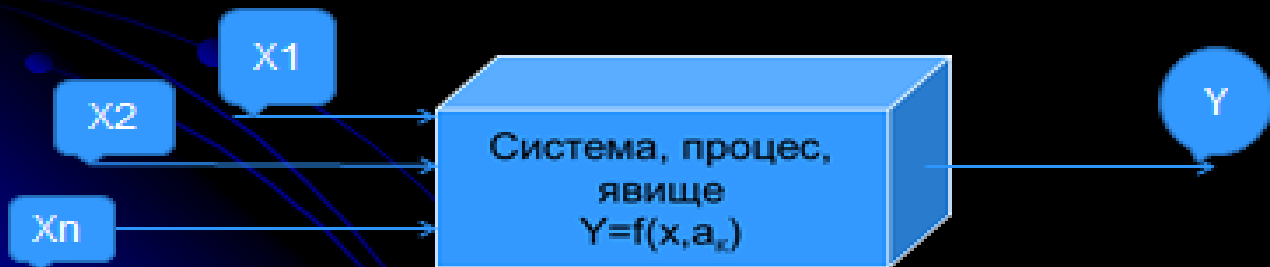
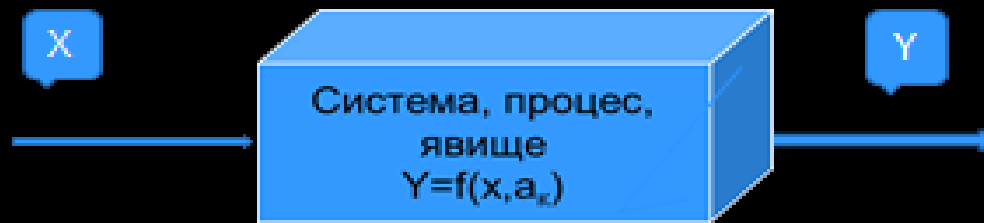
Мета 3:

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



Задачі
апроксимації

Задачі апроксимації

- **О**бробка 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 зовні;
 - задача регресії – побудова найближчої (усередненої) функції
 - задача фільтрації (згладжування) – побудова апроксимуючої функції для зниження систематичної похибки експериментальних даних.
-
- **П**ри інтерполяції доцільно з'єднувати експериментальні точки не ламаною лінією, а згладженою кривою. Тому часто використовують **сплайн – інтерполяцію** (лінійну, квадратичну, кубічну)
-
- **В**сі функції згладжування містять аргументом масиви даних і видають в результаті вектор згладжених даних. Тому, доцільно виконувати згладжування разом з інтерполяцією або регресією.

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.

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

В розгорнутому вигляді можна записати:

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

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 дослід (спостережень) - є метод найменших квадратів (МНК), результатом застосування якого є отримання числових коефіцієнтів емпіричної формули (вибраного багаточлена $P(x)$).

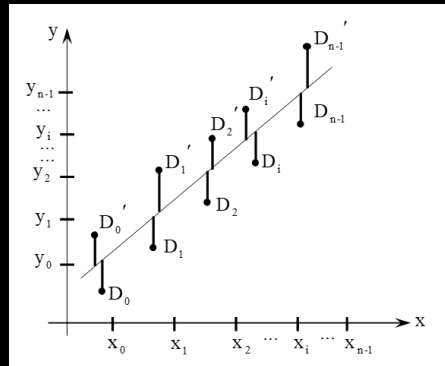
Отже, побудова оптимального апроксимуючого багаточлена $P(x)$ зводиться до знаходження коефіцієнтів, які мінімізують функцію

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

Алгоритм



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).

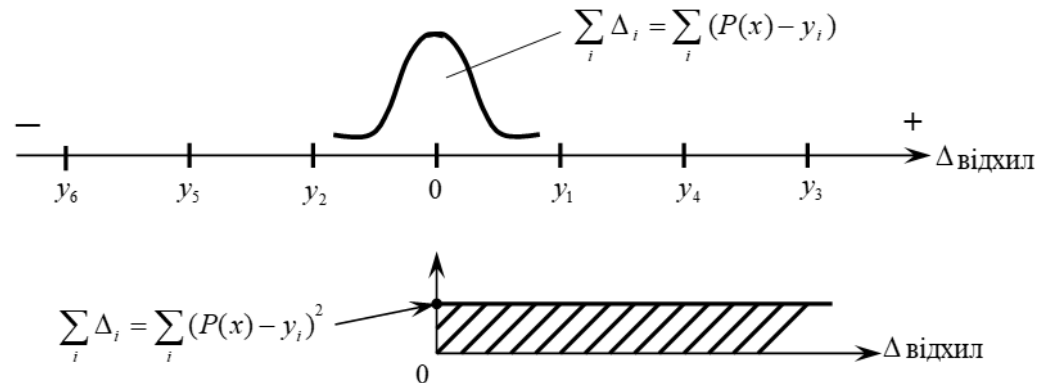


Рис. 2.15

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

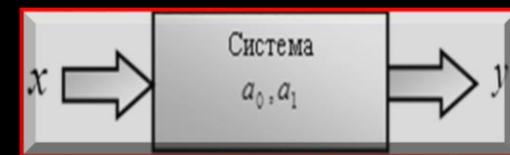
Таким чином, метод апроксимації, суть якого є мінімізація суми квадратів неув'язок $P(x) - y_i$, веде свій початок від праць таких класиків математичної науки, як Лежандр і Гаусс, і називається **методом найменших квадратів**

Розглянемо задачу обчислення коефіцієнтів для лінійної функції

$$P(x_i) = a_0 + a_1 x_i$$

Підставляємо у формулу:

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



Розкриваємо дужки

$$\begin{aligned} [(a_0 + a_1 x_i) - y_i]^2 &= (a_0 + a_1 x_i)^2 - 2(a_0 + a_1 x_i)y_i + y_i^2 = \\ &= a_0^2 + 2a_0 a_1 x_i + a_1^2 x_i^2 - 2a_0 y_i - 2a_1 x_i y_i + y_i^2. \end{aligned}$$

Запишемо систему частинних похідних і прирівняємо їх до нуля:

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

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

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

Розв'язуючи систему відносно a_0 та a_1 , маємо:

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

Методи лінійного згладжування («ковзне середнє»)

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

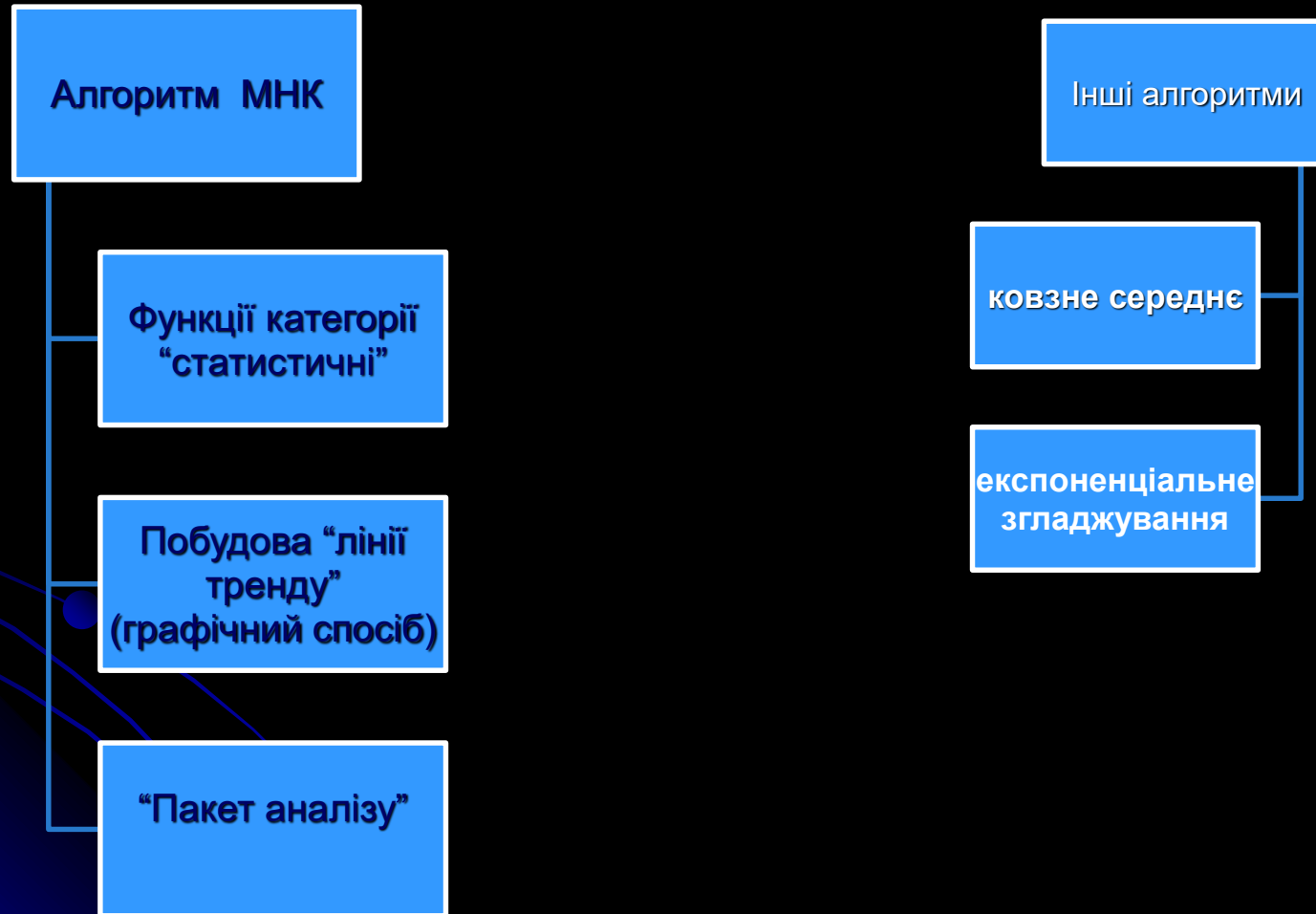
та
експоненціального
згладжування

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

4. Інструментарій апроксимації даних в Excel та Mathcad

4.1. Інструментарій програми Excel

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



Аргументи функції



INTERCEPT

Відомі_значення_y



= масив

Відомі_значення_x



= масив

=

Знаходить точку перетину осі Y із лінією лінійної регресії, обчисленої на основі відомих значень x і y.

Відомі_значення_y залежна множина спостережень або даних – числа, імена, масиви або посилання на числа.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати

Аргументи функції



SLOPE

Відомі_значення_y



= масив

Відомі_значення_x



= масив

=

Повертає нахил лінії лінійної регресії за вказаними точками даних.

Відомі_значення_y масив або діапазон залежних точок даних – числа, імена, масиви або посилання на числа.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати

Аргументи функції?×

RSQ

Відомі_значення_y



= масив

Відомі_значення_x



= масив

=

Повертає квадрат коефіцієнта кореляції Пірсона, обчисленого за вказаними точками.

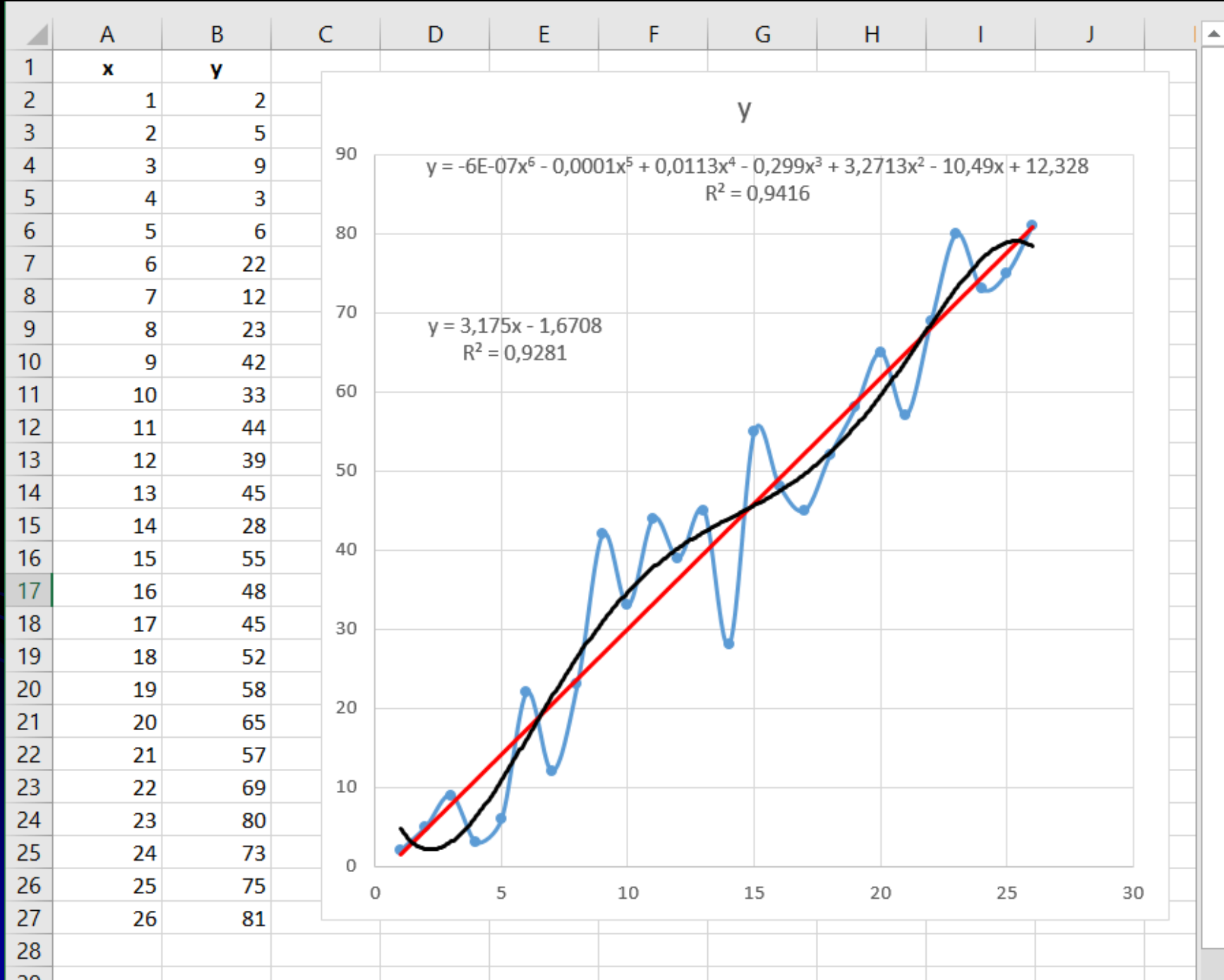
Відомі_значення_y масив або діапазон точок даних – числа, імена, масиви або посилання на числа.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати



Формат лінії тренду

Параметри лінії тренду ▾



▲ Параметри лінії тренду



☐ Експонентна



☒ Лінійна



☐ Логарифмічна



☐ Поліноміальна

Порядок



☐ Степеневий



☐ Змінне
середнє

Період

Назва лінії тренду

☒ Автоматична

Лінійна (y)

☐ Настроювані

Прогноз

Вперед

періодів

Назад

періодів

☐ Установити перетин

☐ Показувати рівняння на діаграмі

☐ Показувати величину вірогідності
апроксимації (R^2) на діаграмі

Аргументи функції



LINEST

Відомі_значення_y			=	посилання
Відомі_значення_x			=	посилання
Конст			=	логічне
Статистика			=	логічне

=

Повертає параметри лінійного наближення, обчислені за методом найменших квадратів.

Відомі_значення_y сукупність значень y, уже відомих у відношенні $y = mx + b$.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати

Аргументи функції



LOGEST

Відомі_значення_y



= посилання

Відомі_значення_x



= посилання

Конст



= логічне

Статистика



= логічне

=

Повертає параметри експоненційного наближення.

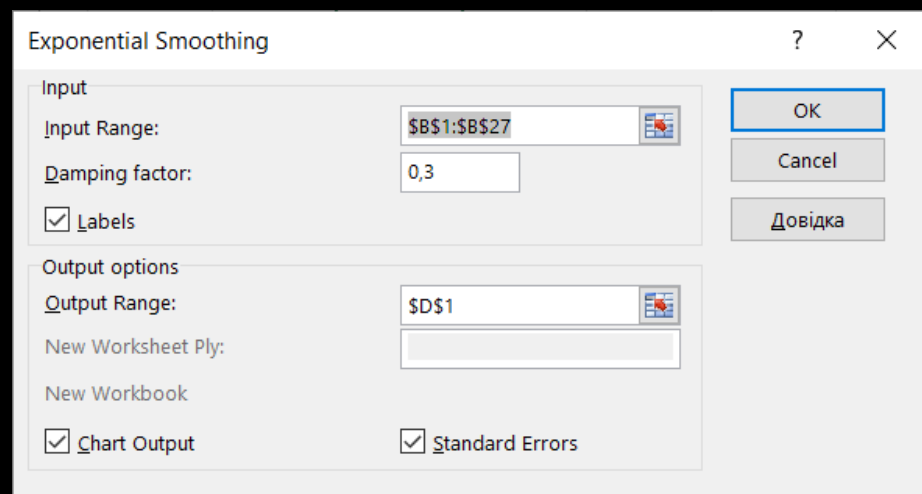
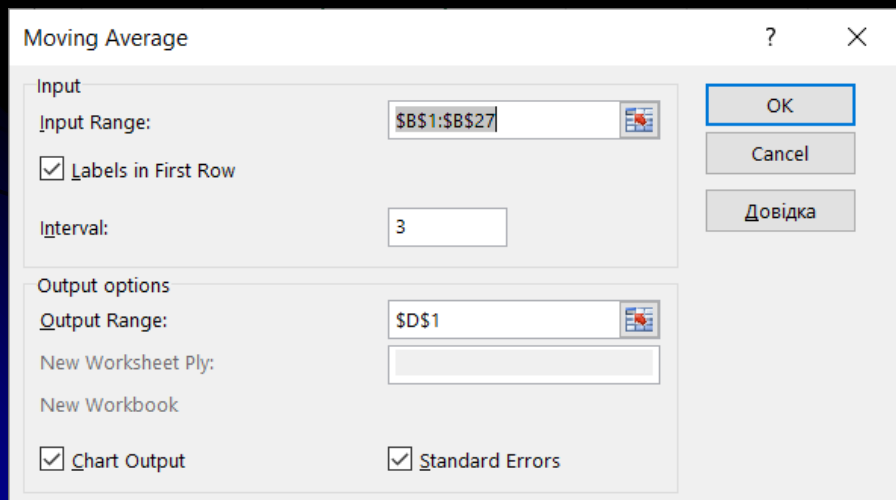
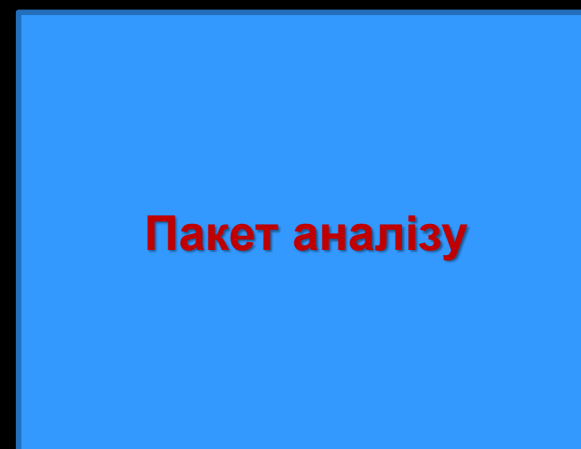
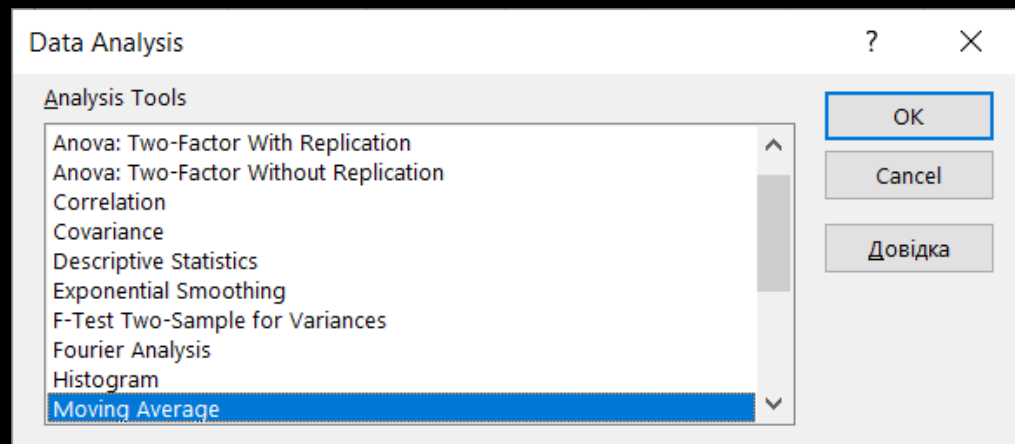
Відомі_значення_y сукупність значень y, уже відомих у відношенні $y = b \cdot m^x$.

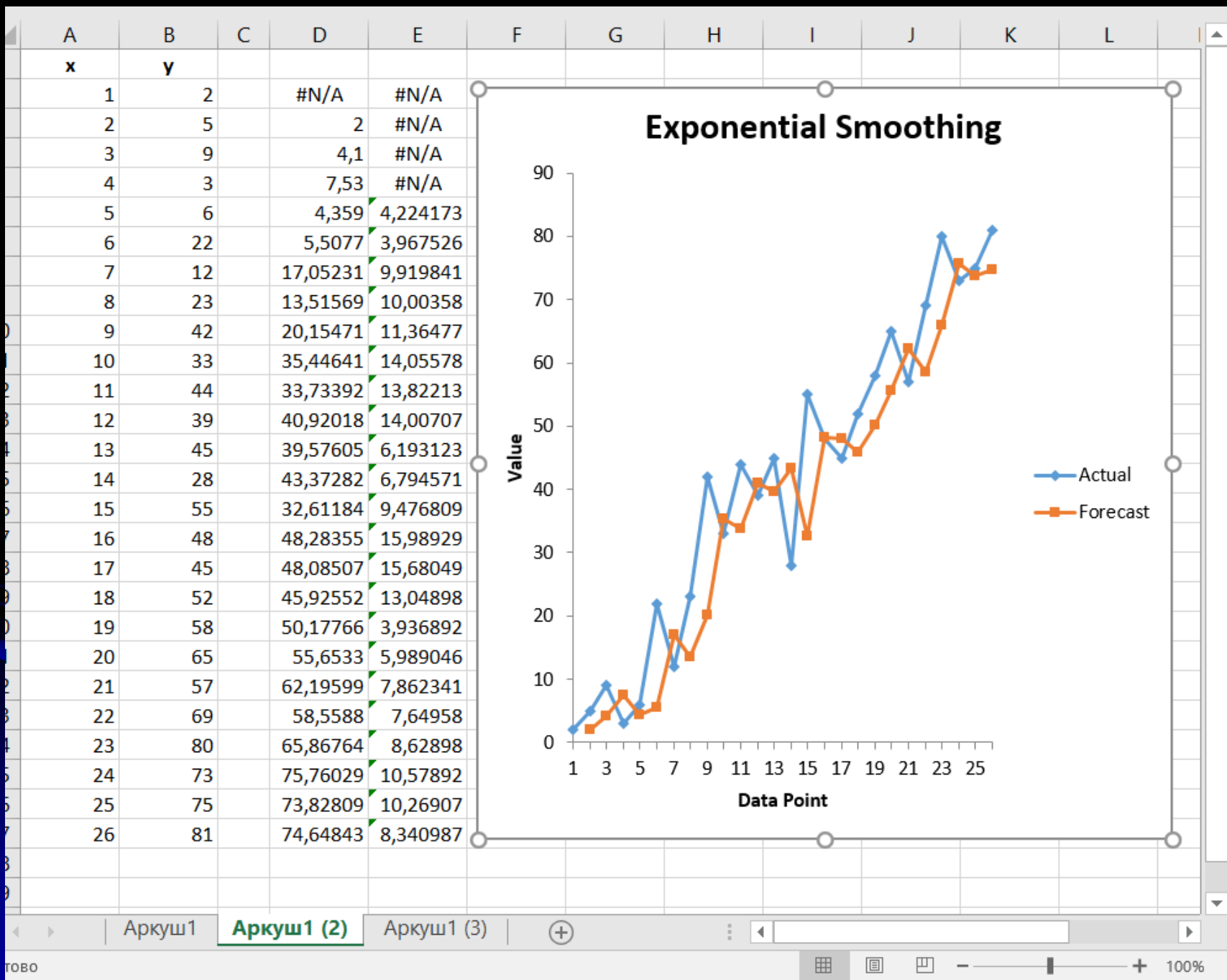
Значення:

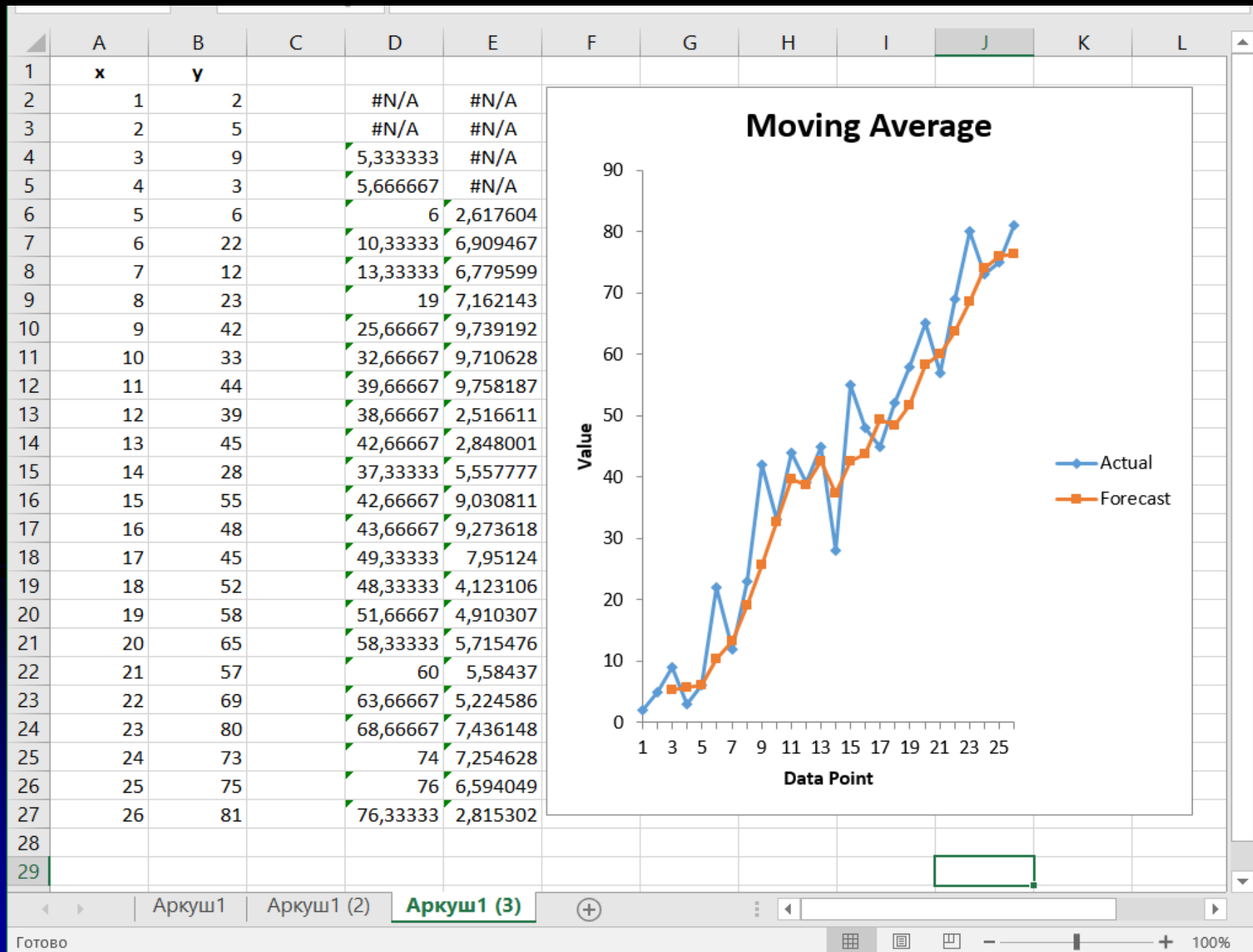
[Довідка з цієї функції](#)

OK

Скасувати







Regression



Input

Input Y Range:



Input X Range:



☐ Labels

☐ Constant is Zero

☐ Confidence Level:

95 %

OK

Cancel

Довідка

Output options

☐ Output Range:



☒ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals

☐ Residual Plots

☐ Standardized Residuals

☐ Line Fit Plots

Normal Probability

☐ Normal Probability Plots

4.1.2. Задача екстраполяції

Алгоритм МНК

Функції категорії "статистичні"
ТЕНДЕНЦІЯ, ПРЕДСКАЗ, РОСТ

Побудова "лінії тренду" (графічний спосіб)

"Пакет аналізу"

"Маркер заповнення" (ЛКМ) – лінійне наближення

"Маркер заповнення" (ЛКМ) – лінійне наближення або
експоненціальне згладжування

- Функції «Тенденція» **TREND** апроксимує прямою лінією, а функція «Рост» **GROWTH** експонентою (за методом найменших квадратів)

$$\begin{aligned} \text{"Тенденция"} &\Rightarrow \bar{y} = a_0 + a_1 * \bar{x}; \\ \text{"Рост"} &\Rightarrow \bar{y} = a_0 * \exp(a_1 \bar{x}). \end{aligned}$$

- Параметри функцій однакові (для обох функцій): **TREND** (значення Y; старі значення X ; нові значення X для прогнозу; логічна константа: якщо «1» - a_0 обчислюється, якщо «0» - $a_0 = 0$).

Аргументи функції



TREND

Відомі_значення_y



= посилання

Відомі_значення_x



= посилання

Нові_значення_x



= посилання

Конст



= логічне

=

Повертає значення відповідно до лінійного наближення, обчисленого методом найменших квадратів на основі відомих даних.

Відомі_значення_y сукупність значень y, уже відомих у відношенні $y = mx + b$.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати

Аргументи функції



GROWTH

Відомі_значення_y			=	посилання
Відомі_значення_x			=	посилання
Нові_значення_x			=	посилання
Конст			=	логічне

=

Повертає значення відповідно до експоненційного наближення, обчисленого на основі відомих даних.

Відомі_значення_y сукупність значень y, уже відомих у відношенні $y = b \cdot m^x$, масив або діапазон додатних чисел.

Значення:

[Довідка з цієї функції](#)

OK

Скасувати

4.2. Інструментарій програми Mathcad

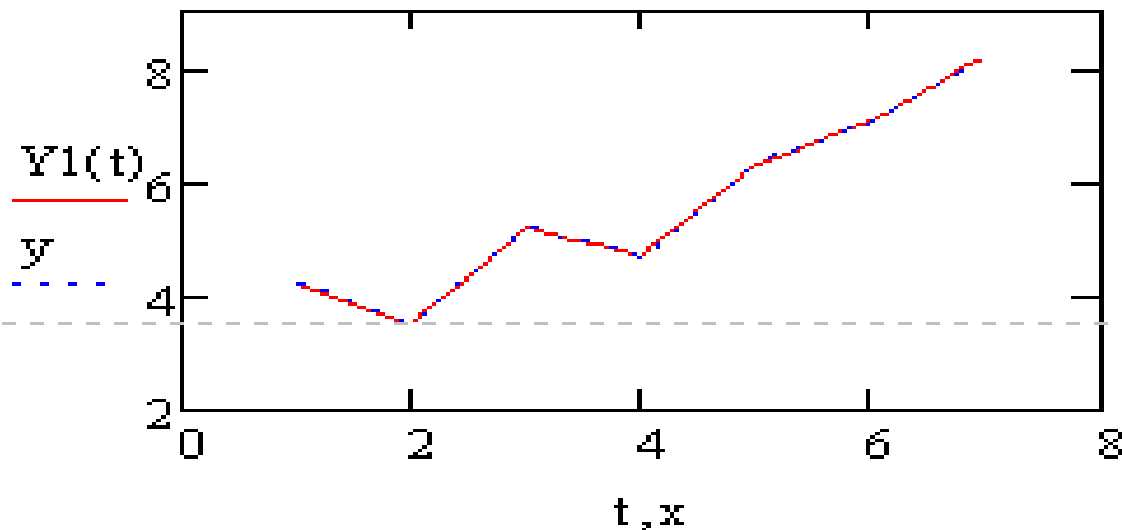
4.2.1. Інтерполяція, екстраполяція, регресія

- **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.

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

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

$$Y1(t) := \text{linterp}(x, y, t) \quad t := 1, 1.2 \dots 7$$



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

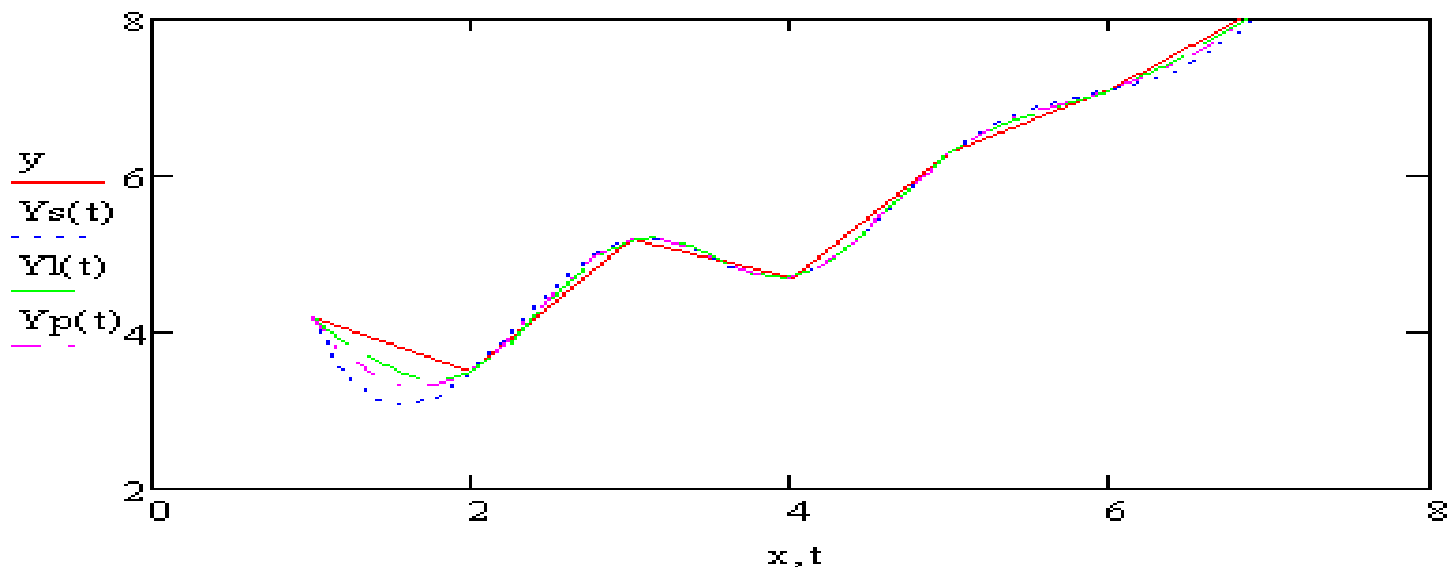
lspline(x,y) - створює вектор значень коефіцієнтів кривої, що наближається до прямої лінії в граничних точках, коефіцієнти лінійного сплайна;

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

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

t - значення аргументу, при якому обчислюється функція

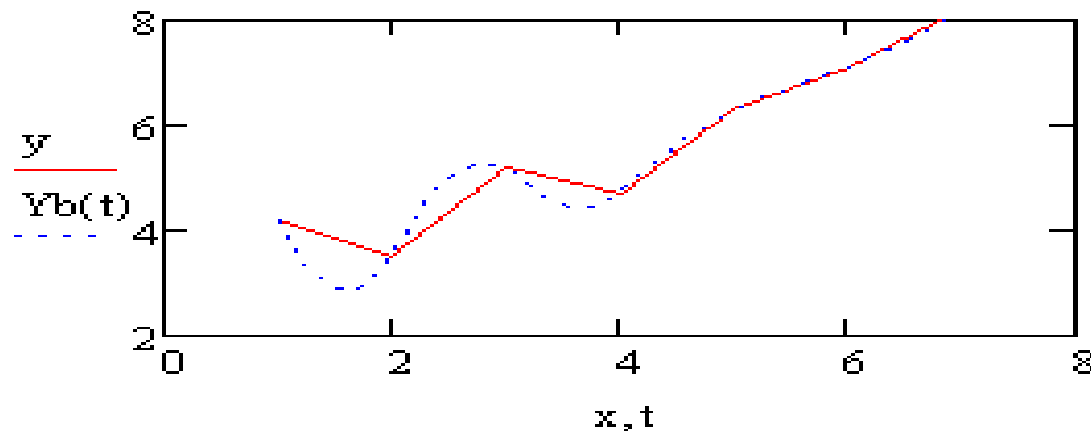
```
x := (1 2 3 4 5 6 7)T  
y := (4.2 3.5 5.2 4.7 6.3 7.1 8.2)T  
s := cspline(x,y)  
Ys(t) := interp(s,x,y,t)  
Yl(t) := interp(lspline(x,y),x,y,t)  
Yp(t) := interp(pspline(x,y),x,y,t)
```



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;

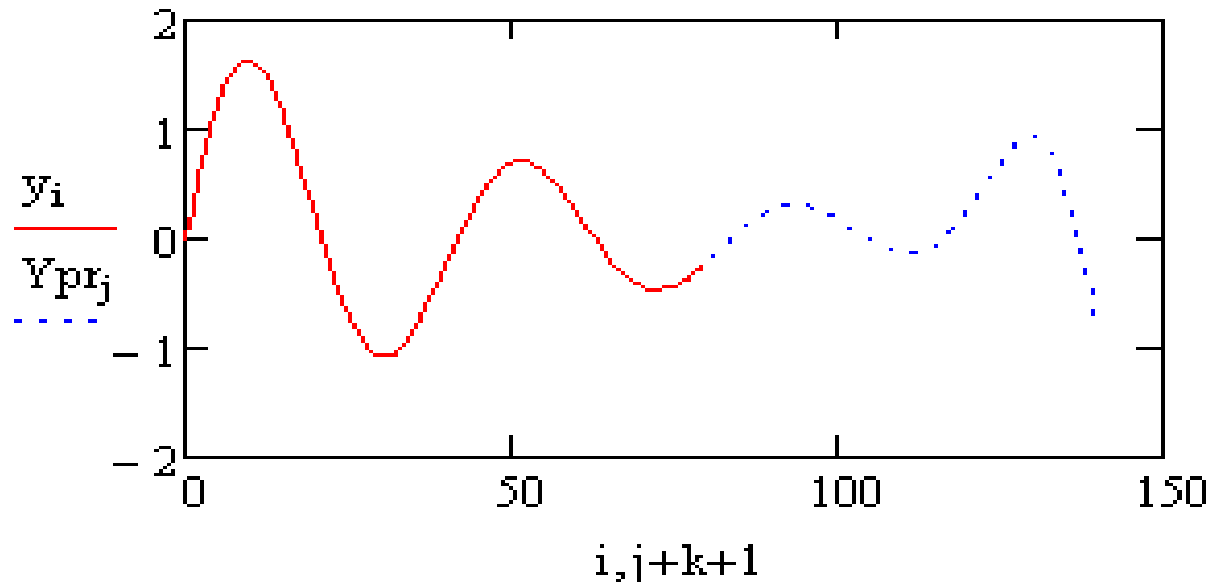
m - кількість послідовних елементів вектора для виконання екстраполяції;

n - кількість точок екстраполяції (прогнозуючих елементів вектора).

```
k := 80    i := 0..k
```

```
yi := 2 · e-0.02·i · sin(0.15·i)
```

```
m := 70    n := 60    Ypr := predict(y,m,n)    j := 0..n
```



Регресія.

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

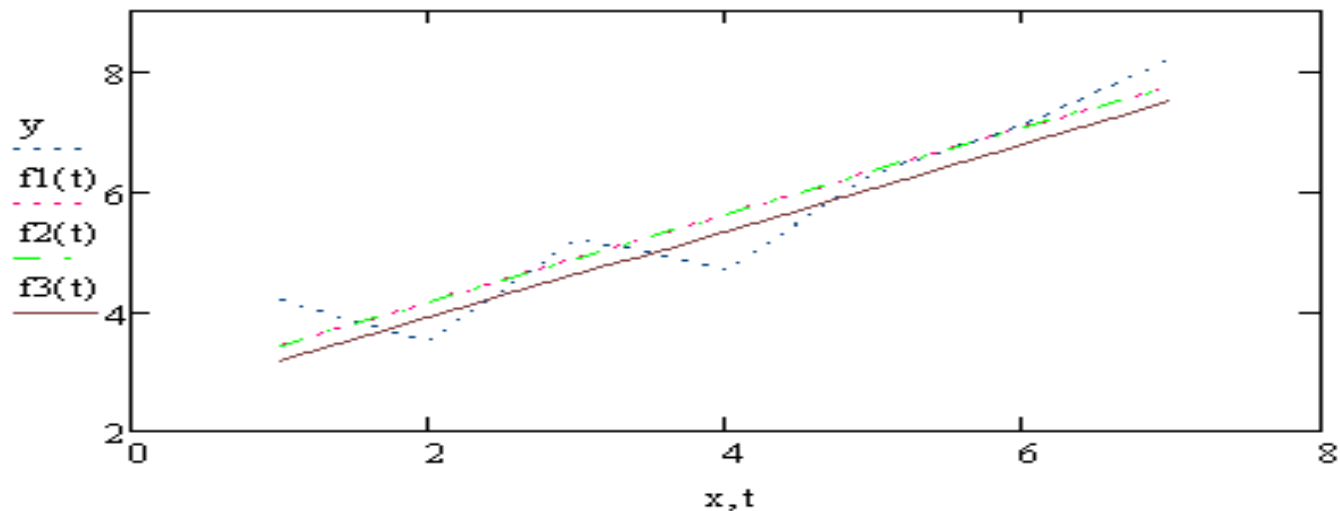
$$x := (1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7)^T$$

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

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

$$\begin{aligned} \text{intercept}(x, y) &= 2.7 & f2(t) &:= \text{intercept}(x, y) + \text{slope}(x, y) \cdot t \\ \text{slope}(x, y) &= 0.725 \end{aligned}$$

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



Поліноміальна регресія

- Спосіб 1. Наближення поліномом k - го степеню ,

$$y(x) = a_0 + a_1x + a_2x^2 + \dots a_kx^k$$

```
x := (1 2 3 4 5 6 7)T
```

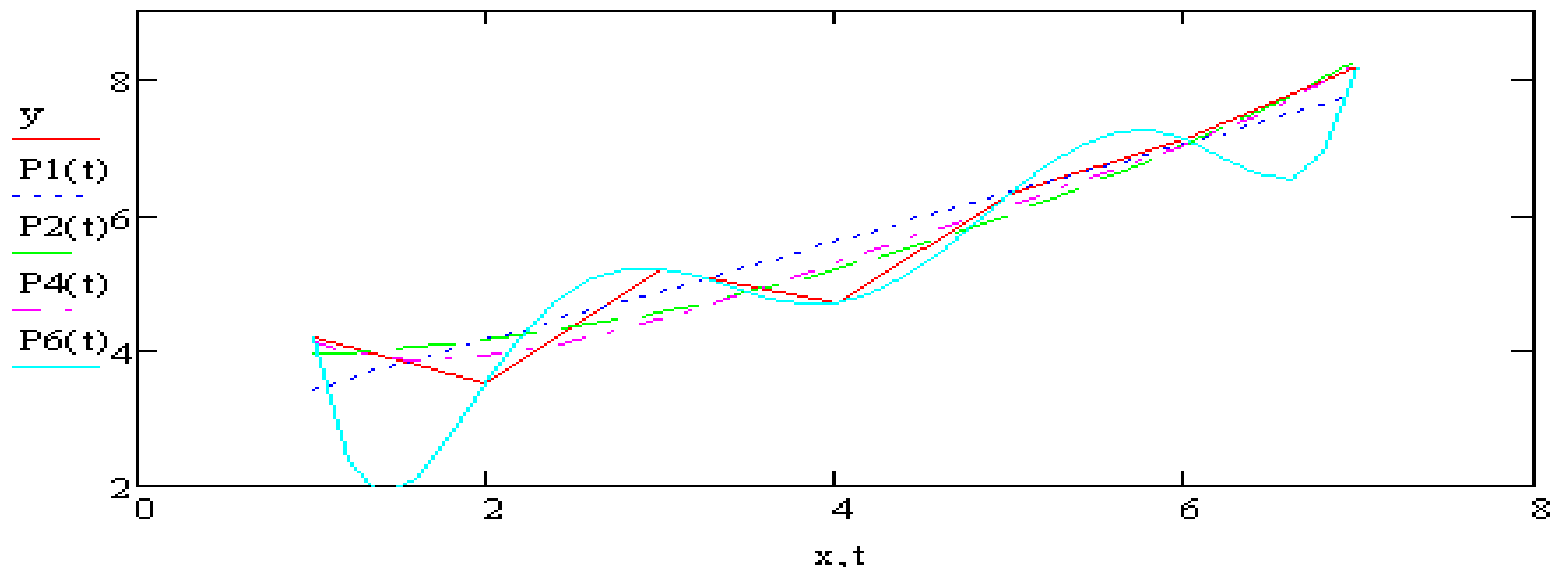
```
y := (4.2 3.5 5.2 4.7 6.3 7.1 8.2)T
```

```
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

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

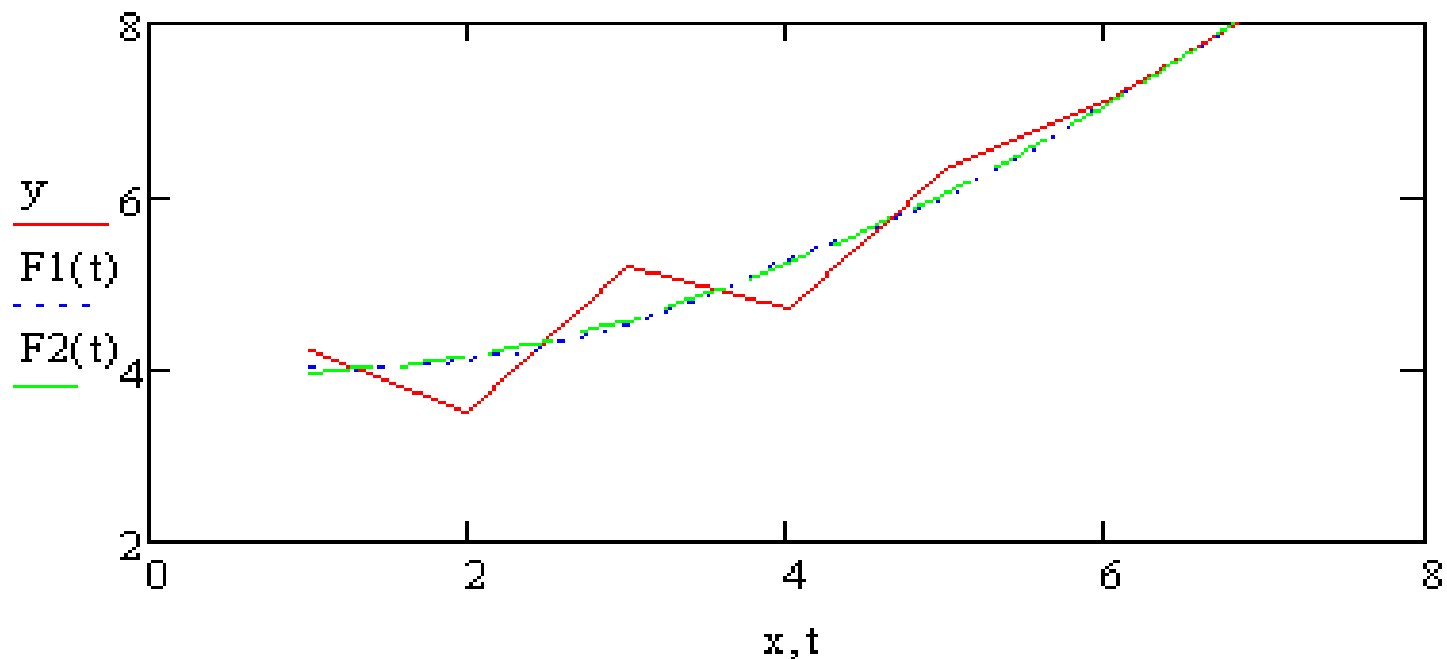
$y := (4.2 \ 3.5 \ 5.2 \ 4.7 \ 6.3 \ 7.1 \ 8.2)^T$

$s1 := \text{loess}(x, y, 0.9)$

$s2 := \text{loess}(x, y, 2)$

$F1(t) := \text{interp}(s1, x, y, t)$

$F2(t) := \text{interp}(s2, x, y, t)$



Регресія на базі нелінійних математичних функцій

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

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

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

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

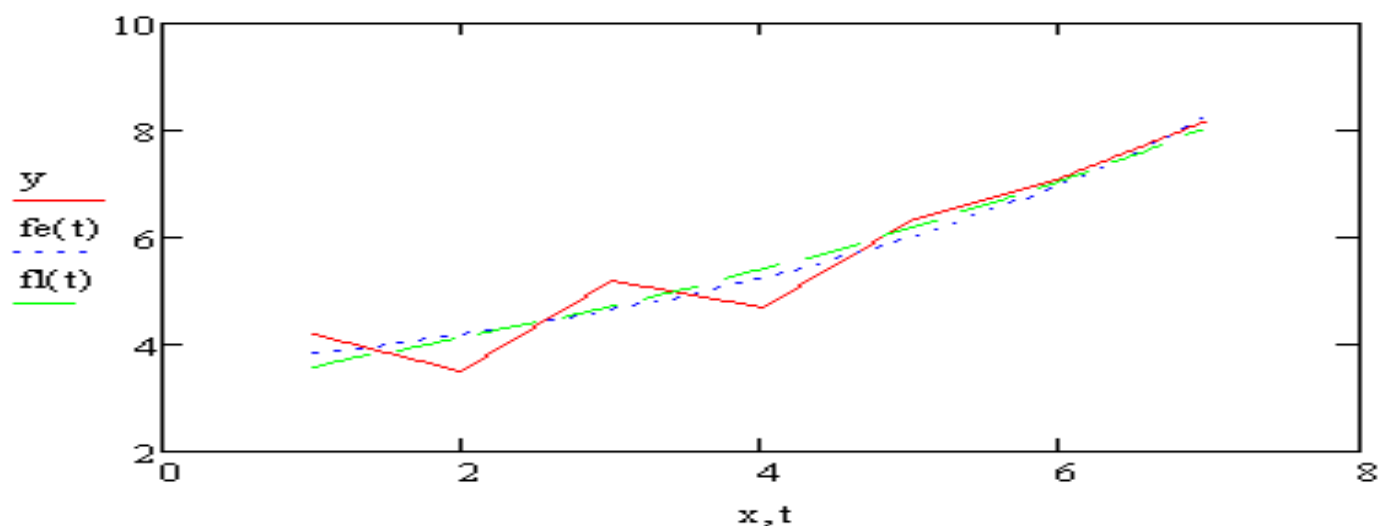
$$C = \begin{pmatrix} 0.917 \\ 0.26 \\ 2.646 \end{pmatrix}$$

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

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

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

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



Регресія загального вигляду

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

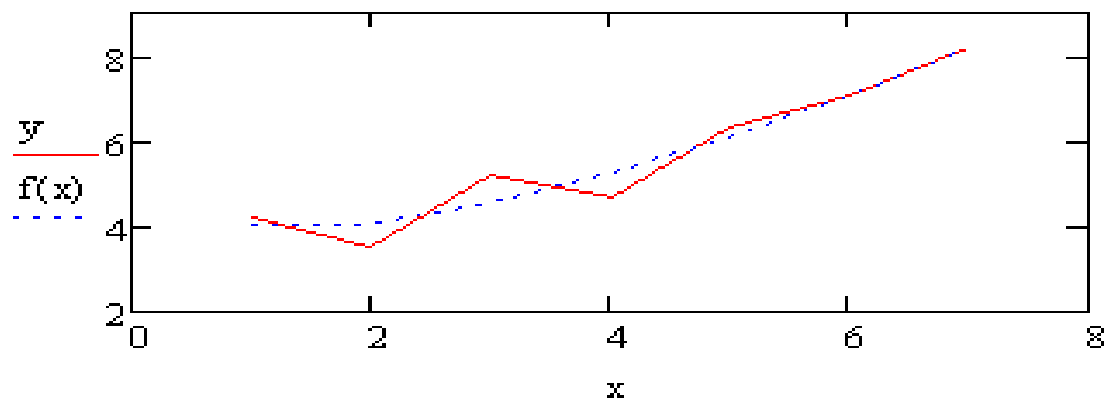
$$y := (4.2 \ 3.5 \ 5.2 \ 4.7 \ 6.3 \ 7.1 \ 8.2)^T$$

$$F(x) := \begin{pmatrix} \frac{1}{x+1} \\ e^{0.2x} \\ x \end{pmatrix}$$

$$C := \text{linfit}(x, y, F)$$

$$C = \begin{pmatrix} 4.882 \\ 0.768 \\ 0.639 \end{pmatrix}$$

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



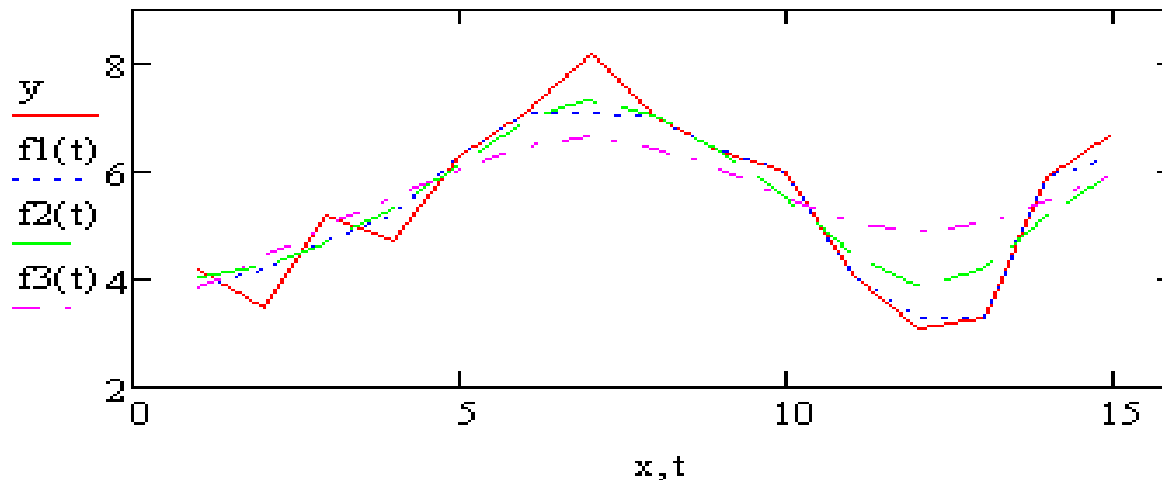
2. Реалізація згладжуючих алгоритмів.

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 найближчої (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.

Додаткові питання

Аналіз даних та елементи імітаційного моделювання

1. *Технологія підбору параметрів*

2. *Таблиця підстановки*

3. *Моделювання випадкових величин*

1. Технологія підбору параметрів.

- Процедура підбору параметрів виконується для прогнозування значень одного параметру при змінюванні іншого параметру.
- Розглянемо сутність процедури на принципі найпростішої моделі росту

$$W = W_0 + b \cdot t$$

- де W – економічний показник
- W_0 – початкове значення
- t – час
- b – параметр функції росту, швидкість зростання
- $f(t) = b \cdot t$ – коефіцієнт зростання.



C26			
	A	B	C
1	Початкова маса W_0 , кг	3	
2	Час t , доба	10	
3	Параметр росту b , кг/добу	0,02	
4	Функція росту $f(t)=b*t$, кг	0,2	
5	Маса $W=W_0+f(t)$, кг	3,2	
6			
7			
8			
9			
10			

?

X

Подбор параметра

Установить в ячейке:

\$B\$5

Значение:

7

Изменяя значение ячейки:

\$B\$2

OK

Отмена

2. Таблиці підстановки.

- Технологія таблиці підстановки призначена для отримання ряду значення результуючої функції, при зміні параметрів, які впливають на результуючу функцію. Розглянемо сутність технології для попереднього прикладу.
- Нехай потрібно вивчити процес змінювання маси об'єкту W , та функції росту $f(t)$, від параметра росту (b), а також іншу задачу – залежність маси об'єкту від параметру росту b та часу t .



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Початкова маса W_0 , кг	3		b	f(t)	W			W	t										
2	Час t, доба	10		0,02	0,2	3,2			3,2	10	20	50	80	120	150	200	250	300	365	450
3	Параметр росту b, кг/добу	0,02		0,03	0,3	3,3			0,02	3,2	3,4	4	4,6	5,4	6	7	8	9	10	12
4	Функція росту f(t)=b*t, кг	0,2		0,04	0,4	3,4			0,03	3,3	3,6	4,5	5,4	6,6	7,5	9	11	12	14	17
5	Маса $W=W_0+f(t)$, кг	3,2		0,07	0,7	3,7			0,04	3,4	3,8	5	6,2	7,8	9	11	13	15	18	21
6				0,09	0,9	3,9			0,07	3,7	4,4	6,5	8,6	11	14	17	21	24	29	35
7				0,1	1	4			0,09	3,9	4,8	7,5	10	14	17	21	26	30	36	44
8				0,15	1,5	4,5		b	0,1	4	5	8	11	15	18	23	28	33	40	48
9				0,18	1,8	4,8			0,15	4,5	6	10,5	15	21	26	33	41	48	58	71
10				0,2	2	5			0,18	4,8	6,6	12	17	25	30	39	48	57	69	84
11				0,25	2,5	5,5			0,2	5	7	13	19	27	33	43	53	63	76	93
12				0,3	3	6			0,25	5,5	8	15,5	23	33	41	53	66	78	94	116
13				0,35	3,5	6,5			0,3	6	9	18	27	39	48	63	78	93	113	138
14				0,4	4	7			0,35	6,5	10	20,5	31	45	56	73	91	108	131	161
15				0,5	5	8			0,4	7	11	23	35	51	63	83	103	123	149	183

3. Моделювання випадкових величин.

Інструментарій Excel для генерування випадкових чисел

- **1. Функція СЛЧИС()** – функція без аргументів категорія «математичні» обчислює випадкові числа, які рівномірно розподілені на інтервали $[0,1]$.
- **Особливість:** формули, які містять функцію СЛЧИС, перераховуються при кожному перерахунку робочого аркуша (наприклад, при введенні значень у комірку або вилученні, або при натисненні клавіші [F9]), що важливо для машинного моделювання. Для фіксації результатуючих значень функції СЛЧИС, можна їх скопіювати «Спеціальною вставкою».
- **2. Функція СЛУЧМЕЖДУ** (нижня границя; верхня границя) – виконує генерацію цілочисленних значень, підпорядкованих дискретному рівномірному розподілу. Функція знаходиться в категорії «математичні та тригонометричні» і доступна тільки після підключення надбудови «Пакет аналізу».
- **Особливості:**
 - як і функція СЛЧИС, при кожному перерахунку функція СЛУЧМЕЖДУ змінює значення;
 - функція «не працює» у формулах масивів;
 - якщо значення аргументів дробні, то вони округлюються до найближчих цілих.

3. Програма «Генерація випадкових чисел» із надбудови «Пакет аналізу» призначений для отримання випадкових вибірок, тобто генерації значень випадкових чисел заданого розподілу: рівномірного, нормального, Бернуллі, біноміального, Пуассона, модельного та дискретного.

Деякі технології використання функції «СЛЧИС»:

- 1). Моделювання випадкових величин, які рівномірно розподілені на довільному інтервалі , виконується формулою:

$$= (b - a)^* \text{СЛЧИС}() + a.$$

- 2). Генерування цілочисельних значень, які рівномірно розподілені на інтервалі - аналог функції «СЛУЧМЕЖДУ», виконується формулою:

$$= \text{ЦЕЛОЕ}((b - a)^* \text{СЛЧИС}() + a).$$

- 3). Використання перерахунку функції «СЛЧИС» для отримання результатів декількох експериментів, тобто значення характеристик, наприклад: середнє, дисперсія, мінімум, максимум тощо. Для цього можна використати «Таблицю підстановки»

Технологія:

- за допомогою {= СЛЧИС ()} за формулою масиву отримати вибірку значень (стовпчик А);
- стандартними функціями обчислити характеристики: середнє, дисперсію, мінімум, максимум (стовпчик В);
- ввести номери експериментів 1, 2, 3, ... N, наприклад у стовпчик D;
- у стовпчики Е, F, G, Н ввести формули із посиланнями до комірок з відповідними характеристиками;
- виділити діапазон для «Таблиці підстановки» (D3: Н10);
- меню «Дані», «Таблиця підстановки»;
- в полі «Підставляти значення по рядках» ввести посилання до будь-якої вільної комірки (наприклад, К 2), оскільки числа, які позначають кількість експериментів (стовпчик D), в обчисленнях участі не приймають, а слугують «пусковим механізмом» для нового перерахунку формул, які містять функцію «СЛЧИС»;
- кнопка «ОК», після чого, таблиця буде заповнена значеннями відповідних характеристик для кожного експеримента.

	А	В	С	Д	Е	F	G	Н
1	вибірка	середнє						
2	0,991337	=СРЗНАЧ(А2:А10)		№ експеримента	сер	дисп	min	max
3	0,794601	дисп			=В2	=В4	=В6	=В8
4	...	=ДИСП(А2:А10)		1				
5	...	min		2				
6	...	=МИН(А2:А10)		3				
7	...	max		4				
8	...	=МАКС(А2:А10)		5				
9	0,128896			6				
10	0,103461			7				
11				8				
12				9				