Національний технічний університет України «КПІ ім. Ігоря Сікорського» Факультет Інформатики та Обчислювальної Техніки

Кафедра інформаційних систем та технологій

Лабораторна робота №1

з дисципліни «Інтелектуальний Аналіз Даних»

на тему

«МЕТОДИ ОБРОБКИ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ»

Варіант №12

Виконав:

студент групи ІС-02

Плостак І. М.

Київ – 2021

**1.1. Мета роботи**

Сформувати у студентів уявлення про підхід до вирішення завдання про середньоквадратичне наближення функції, заданої таблично; прищепити знання про методи апроксимації елементарними функціями; виробити навички роботи в середовищі МАТLAB.

**1.2. Завдання до роботи**

1. Ознайомитися з конспектом лекцій та рекомендованою літературою за темою роботи, а також з додатком Ж, що містить опис програмного забезпечення для побудови різних видів регресій.
2. Використовуючи дані таблиці і застосовуючи стандартні заміни змінних, знайти рівняння наступних видів регресій: лінійної, гіперболічної, степеневої, показникової, логарифмічної.
3. Порівняти якість отриманих наближень шляхом порівняння їх відхилень.
4. Побудувати графіки одержаних залежностей і табличних значень аргументів і функції.

**2. Короткі теоретичні відомості**

**Види рівнянь регресії. Їх інтерпретація.**

Рівняння, яке описує теоретичну лінію регресії називають **рівнянням регресії**.

Де *f(x)*-якась невідома функція, а х - середня величина ознаки, яка змінюється.

*f(x)* - Функція, яка встановлює вид однозначної залежності між цими величинами - це розрахункові теоретичні значення.

**Рівняння регресії** виражає середню величину однієї ознаки як функцію іншого.

**Функція регресії** - Це модель виду залежності у від х, де у – залежна змінна (результативна ознака); х - незалежна, або пояснює, змінна (ознака-фактор).

**Лінія регресії** - Графік функції у=f(x).

2 типи взаємозв'язків між х і у:

1. може бути невідомо, яка з двох змінних є незалежною, а яка - залежною, змінні рівноправні, це взаємозв'язок кореляційного типу;
2. якщо х і у нерівноправні і одна з них розглядається як пояснює (незалежна) змінна, а інша - як залежна, то це взаємозв'язок регресійного типу.

Види регресій:

1. Гіперболічна - Регресія рівносторонній гіперболи:

у = а + b / х + Е;

1. Лінійна - регресія, що застосовується в статистиці у вигляді чіткої

економічної інтерпретації її параметрів:

у = а + b \* х + Е;

1. Логарифмічно лінійна - регресія виду:

ln у = ln а + b \* ln x + ln E

1. Множинна - регресія між змінними у та х1 , х2 ...xm,

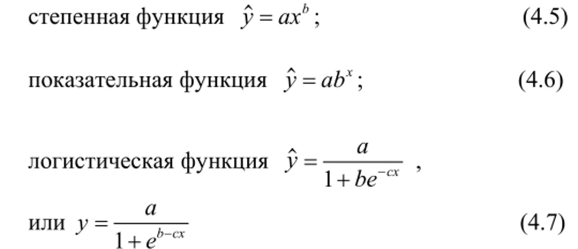
Модель виду: у = f (х1 , х2 ...xm) + E, де у - залежна змінна (результативна ознака), х1, х2 ...xm - Незалежні, пояснюючі змінні (ознаки-фактори), Е - збурення або стохастична змінна, що включає вплив неврахованих факторів в моделі

1. Нелінійна - регресія, нелінійна щодо включених в аналіз пояснюють

змінних, але лінійна по оцінюваним параметрам; або регресія, нелінійна по оцінюваним параметрам.

1. Зворотна - регресія, що приводиться до лінійного вигляду, реалізована в стандартних пакетах прикладних програм виду: у = 1 / a + b \* х + Е;
2. Парна - регресія між двома змінними у та x, модель виду: у = f (x) + Е, де у-залежна змінна (результативна ознака), x - незалежна, яка пояснює змінна (ознака - фактор), Е - збурення , або стохастична змінна, що включає вплив неврахованих факторів в моделі.

Істотно нелінійними регресіями є такі функції, найбільш часто використовувані для опису економічних процесів:



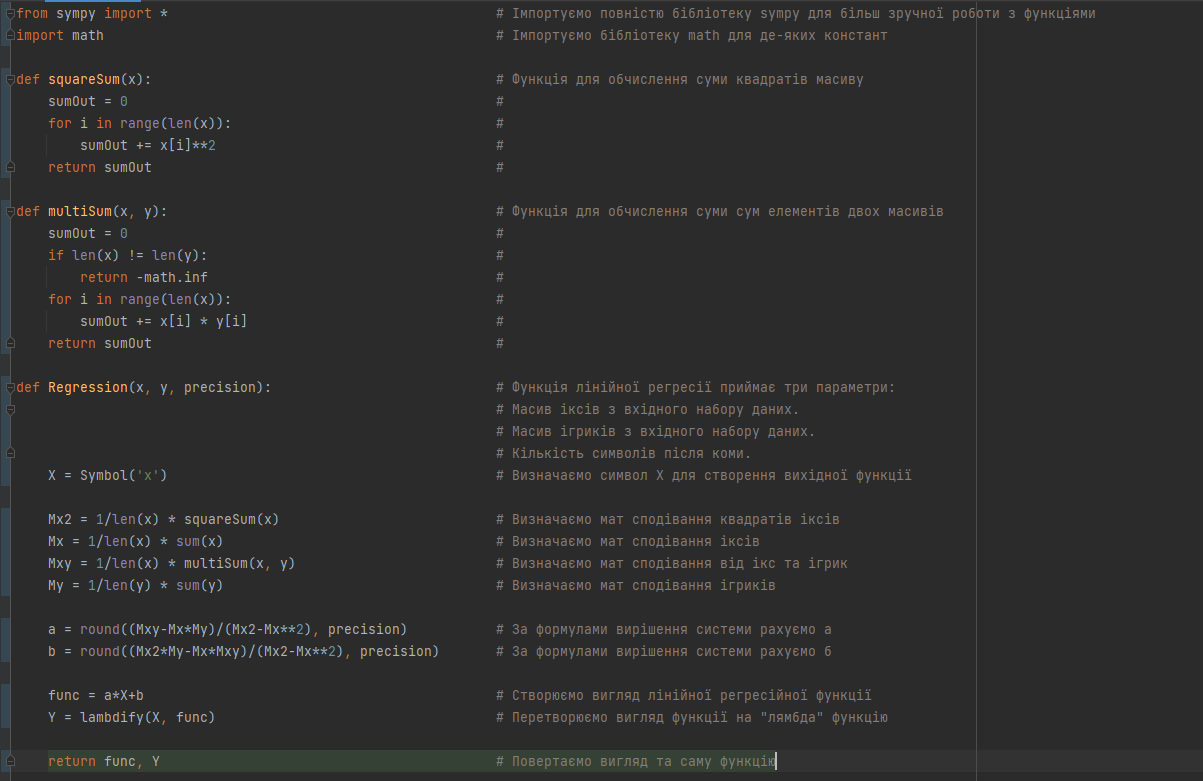
**3. Набір даних для обробки**

Варіант 12:

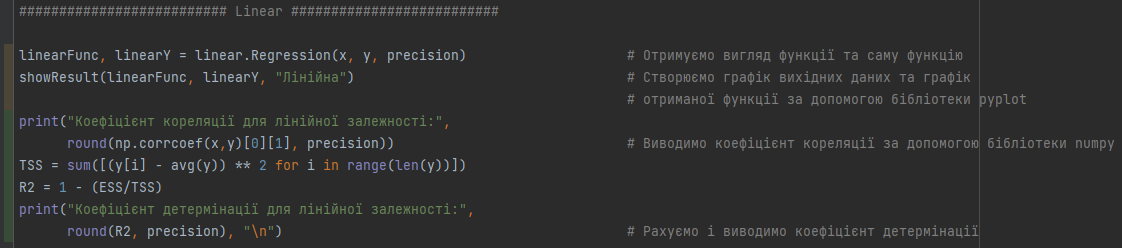
* x = [0.28, 0.19, 0.15, 0.11, 0.09, 0.08, 0.07, 0.06]
* y = [82.16, 61.02, 44.56, 82.52, 99.19, 70.24, 63.23, 66.48]

**4. Детальний опис процесу використання ПЗ**

Детально розглянемо роботу програми на прикладі функції створення вигляду функції лінійної регресії:



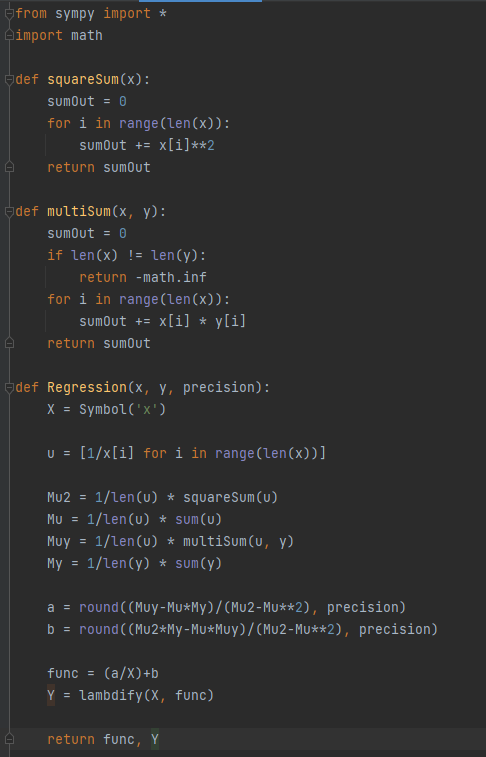
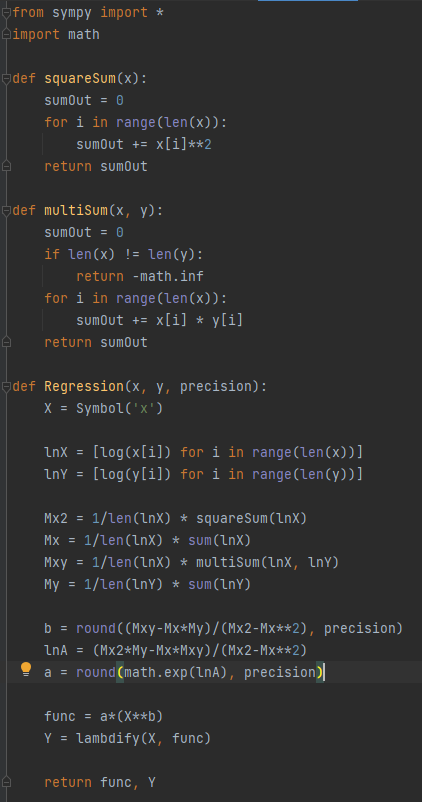




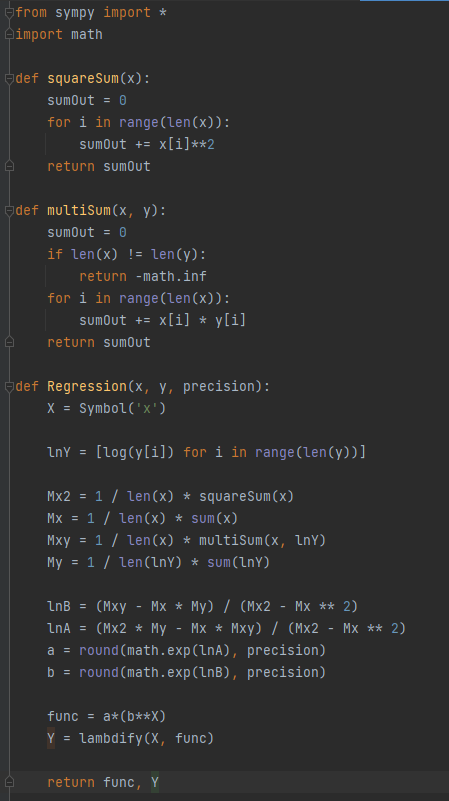
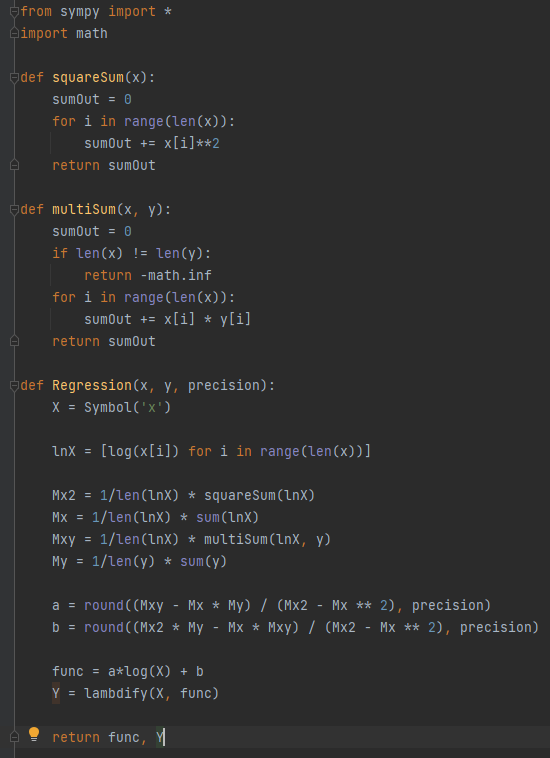
Інші 4 функції виглядають так само, оскільки ми кожну з них зводимо до лінійної регресії за допомогою заміни. Наприклад степенева: y = axb. Логарифмуємо та отримуємо: ln(y) = ln(axb), ln(y) = ln(a) + b\*ln(x) – потрібний нам лінійний вигляд. Робимо заміну ln(y) = u, ln(a) = q, ln(x) = v. Отримали лінійний вигляд: u = q + b\*v. Звідси знаходимо усе необхідне та повертаємося до початкового коефіцієнту а. Аналогічно з іншими трьома регресіями. Код нижче.

**5. Текст програми (інші 4 функції)**

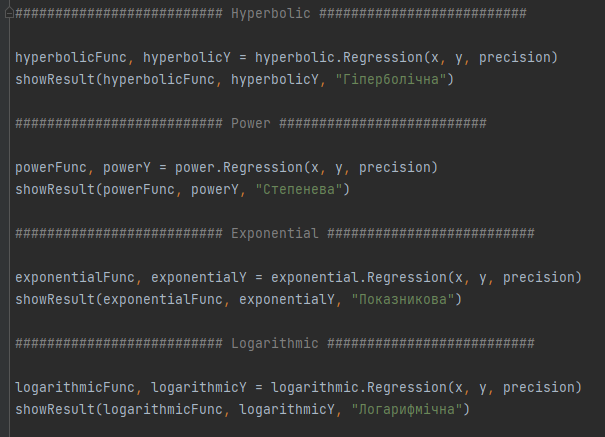
Гіперболічна та степенева:

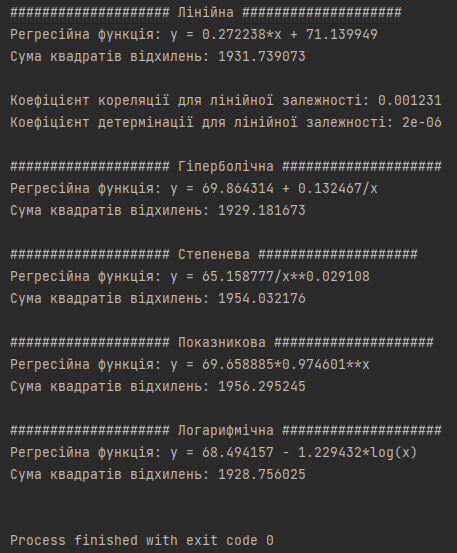
Показникова та логарифмічна:

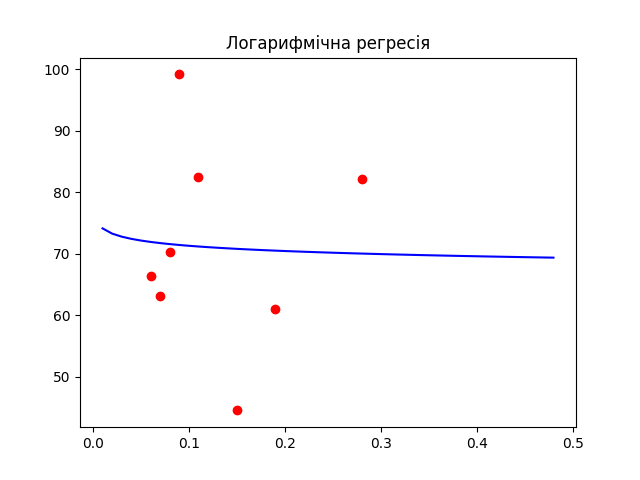
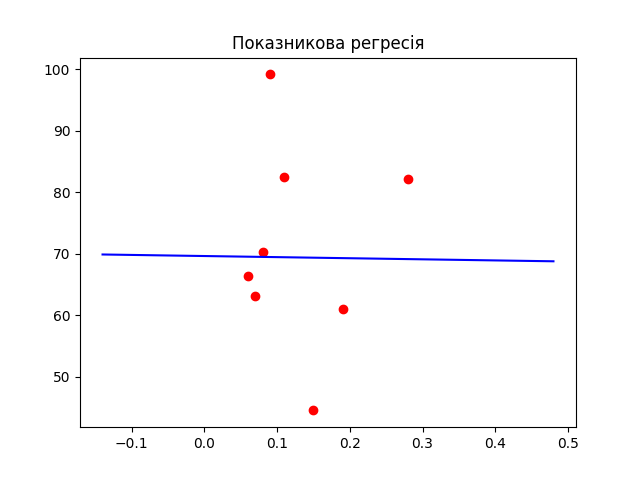
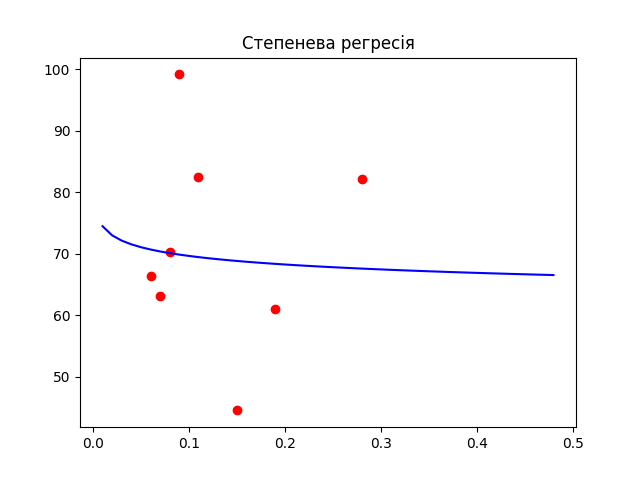
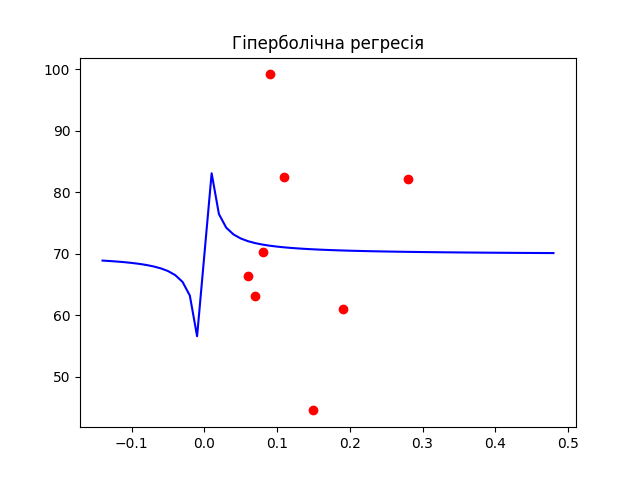
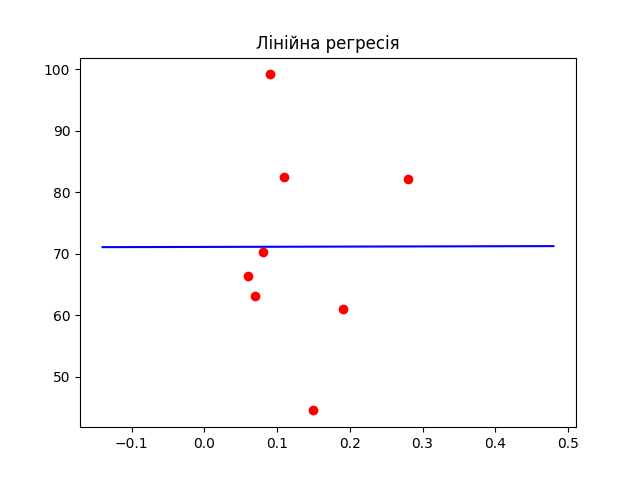
 

Виклик цих функцій:



**6. Результати роботи програмного забезпечення**





**7.1. Контрольні питання**

1. Загальна постановка задачі знаходження наближуючої функції.

*Регресійний аналіз – кількісне уявлення зв’язку або залежності між залежною змінною (відгуком) і незалежної / незалежними змінними(предикторами).*

*Регресійний аналіз використовується з двох причин:*

*1) опис залежності між змінними допомагає встановити наявність можливого причинного зв’язку;*

*2) для встановлення предиктора для залежної змінної, так як рівняння регрессії дозволяє прогнозувати значения залежної змінної за значеннями незалежних змінних: виявлення закономірності, яка зображена у вигляді рівняння регресії.*

1. У чому суть наближення таблично заданої функції за методом найменших квадратів?

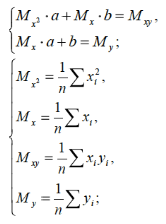
*Задається вигляд регресійної моделі і шукаються параметри моделі, які мінімізують суму квадратів відхилень.*

1. Які функції можуть бути використані в якості наближаючих?

*Лінійна, степенева, гіперболічна, логарифмічна, показникова тощо.*

1. Як знаходяться відхилення виміряних значень Y від обчислених за формулою наближуючої функції?
2. Як знайти наближуючу функцію у вигляді лінійної функції F(x,a,b)=ax+b?

*Обчислити коефіцієнти a і b рівняння лінійної регресії можна, скориставшись системою:*



1. Як знайти наближуючу функцію у вигляді квадратичної функції F(x,a,b,c)=ax^2 +bx+c?

*У нас є система з n рівнянь, де n – кількість вимірів з трьома невідомими a, b, c:*

, де X = , Y =.

*Отже, рішенням такої системи буде:*

1. Як привести показникову, степеневу, логарифмічну функції до лінійної?

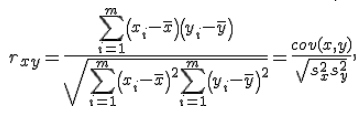
*За допомогою заміни. Наприклад степенева: y = axb. Логарифмуємо та отримуємо: ln(y) = ln(axb), ln(y) = ln(a) + b\*ln(x) – потрібний нам лінійний вигляд. Робимо заміну ln(y) = u, ln(a) = q, ln(x) = v. Отримали лінійний вигляд: u = q + b\*v. Звідси знаходимо усе необхідне та повертаємося до початкового коефіцієнту а. Аналогічно з іншими трьома регресіями.*

1. Як функція трьох змінних може приймати найменше значення?

*Необхідними умовами мінімуму функції є рівність частинних похідних нулю.*

1. Що таке коефіцієнт кореляції і як він знаходиться?

*Коефіцієнт кореляції характеризує існування лінійної залежності між двома величинами. Знаходиться він за формулами:*



*де , – вибіркові середні xm та ym, sx2, sy2 – вибіркові дисперсії.*

1. Які межі значення коефіцієнта кореляції і що вони показують?

*Коефіцієнт кореляції називають також тіснотою лінійного зв'язку:*

*rxy ∈ [-1,1].*

*| rxy | = 1 – лінійно залежні.*

*| rxy | = 0 – лінійно незалежні.*

1. Що таке відхилення?

*Відхилення – це різниця початкового значення y та значення отриманої функції в цій точці. Знаходиться за формулою:*

e = =

1. Як можна визначити правильність виду обраної функції?

*Порівняти суми квадратів відхилень – та функція, у якої ця сума буде меншою і буде найбільш правильною.*

**7.2. Висновки**

В результаті роботи програми та проведеного аналізу було встановлено, що у вхідному наборі точок між x та y практично немає лінійної залежності, оскільки коефіцієнт кореляції близький до нуля (чисельно дорівнює rxy= 0.001231). Цей же висновок підтверджує і обрахований коефіцієнт детермінації R2 після отримання рівняння лінійної регресії, який також близький до нуля (чисельно дорівнює R2=2\*10-6). Це не означає що між змінними зовсім немає залежності, але інші перевірені мною моделі (гіперболічна, степенева, показникова та логарифмічна) не виявилися набагато кращі ніж лінійна регресія. Про це свідчить невелика різниця сум квадратів відхилень між цими моделями. Незначно кращою порівняно з іншими моделями виявилася логарифмічна з сумою квадратів відхилень S=1928.756.