МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ФРАНКА
ФАКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ ТА
ІНФОРМАТИКИ

Індивідуальне завдання №3 з курсу "Теорія ймовірності та математична статистика"

Виконала:
Студентка групи ПМІ-23
Шувар Софія

Кафедра
ДИСКРЕТНОГО АНАЛІЗУ

Викладач:
Квасниця Галина
Андріївна

17 травня 2021 р.

1 Постановка задачі.

1. За даними кореляційної таблиці обчислити умовні середні y_{xi} $(i = 1, ..., k)$.

2. Побудувати поле кореляції, тобто нанести точки $M_i(x_i; \bar{y}_{xi})$, $i = 1, ..., k$, на координатну площину. На основі цього зробити припущення про вигляд функції регресії (парабола, гіпербола і т.д.)

3. В залежності від вигляду функції регресії скласти відповідну систему рівнянь. Розв'язати її і знайти невідомі параметри вибраної функції регресії.

4. Записати рівняння кривої регресії Y на X : $\bar{y}_x = f(x)$ (з конкретною знайденою в пункті 3 функцією регресії $f(x)$) та побудувати її графік.

5. Обчислити дисперсію величини Y відносно кривоїрегресії Y на X .

6. Визначити суму квадратів відхилень δ^2 умовних середніх від значень функції регресії за формулою.

2 Короткі теоретичні відомості.

Зв'язок між випадковими величинами, зміни яких викликані як спільними випадковими факторами, так і дією інших неоднакових для обох випадкових величин факторів, називають стохастичним.

Найважливішим стохастичним зв'язком є зв'язок, який встановлює залежність між значеннями випадкової величини η і умовним математичним сподіванням $M(\eta|\xi = x)$ випадкової величини η : $M(\eta|\xi = x) = \bar{y}(x)$ або між значеннями випадкової величини ξ і умовним математичним сподіванням $M(\xi|\eta = y)$ випадкової величини ξ : $M(\xi|\eta = x) = \bar{x}(y)$. Залежності такого роду називають регресійними; функції $\bar{y}(x)$ і $\bar{x}(y)$, які визначають такого роду залежності, називають функціями регресії η на ξ і ξ на η, а графіки функцій $\bar{y}(x)$ і $\bar{x}(y)$- лініями регресії η на ξ і ξ на η.

За формою розрізняють лінійну і нелінійну регресії, тобто регресії виражені лінійними і нелінійними функціями.

Теорія криволінійної кореляції розв'язує ті самі задачі, що і теорія лінійної кореляції, а саме:

1. За даними кореляційної таблиці встановлюють форму кореляційного зв'язку, тобто визначають вигляд функції $\bar{y}(x)$ або $\bar{x}(y)$;

2. Оцінюють щільність кореляційного зв'язку, тобто дають оцінку ступеню розсіювання значень випадкової величини Y навколо побудованої кривої регресії $\bar{y}(x)$ (або значень випадкової величини X навколо $\bar{x}(y)$.

Для аналізу кореляції між двома величинами ξ і η проводять n незалежних спостережень, результатами яких є пари чисел $(x_i;\ y_j)$ $(\overline{i = 1, k})$, $(\overline{j = 1, l})$. На основі цих спостережень отримали вибірку, де складова X набула значень $x_1, x_2, ..., x_k$, складова $Y - y_1, y_2, ..., y_l$, а подія $X = x_i, Y = y_j$ мала частоту появи n_{ij} $(\overline{i = 1, k})$, $(\overline{j = 1, l})$. Результати цих спостережень записують у вигляді кореляційної таблиці. За даними кореляційної таблиці обчислюють умовні середні $\bar{y}_{xi}$ $(\overline{i=1,k})$, та $\bar{x}_{yj}$ $(\overline{i=1,k})$, та складають таблиці відповідних умовних середніх.

Для визначення вигляду функції регресії будують точки $(x;\bar{y}_x)$ (або $(y;\bar{x}_y)$) і за їх розміщенням роблять висновок про приблизний вигляд функції регресії. Якщо графік регресії $\bar{y}(x)$ або $\bar{x}(y)$ зображається кривою лінією, то кореляцію називають нелінійною (криволінійною).
Деякі нелінійні кореляції:

1. **Параболічна кореляція.**
   У прямокутній системі координат позначимо всі точки, які відповідають парам чисел $(x_i;y_{xi})$, тобто побудуємо поле кореляції.Припустимо, що точки $M_i(\bar{x}_i;y_{xi})$, $i=1,...,k$, розташовані приблизно на параболі другого порядку. Рівняння параболи – параболічної регресії Y на X будемо шукати у вигляді
   $$f(x) = ax^2 + bx + c, \tag{1}$$
   де a, b, c – невідомі параметри. Із всіх парабол такого виду шукана найближче розташована до точок $M_1, M_2, ..., M_k$, причому точка M_i вибирається n_i разів, $i = 1, ..., k$.

2. **Гіперболічна кореляція.**
   Припустимо, що аналіз залежності між змінними X і Y, вираженої кореляційною таблицею, приводить до вибору форми кореляційної залежності Y на X у вигляді рівняння гіперболи
   $$\bar{y}_x = \frac{a}{x} + b, \tag{2}$$
   а у випадку регресії X на Y –гіперболи
   $$\bar{x}_y = \frac{c}{y} + d. \tag{3}$$
   Регресії такого типу називаються гіперболічними.

3. **Показникова кореляція.**
   Розглянемо випадок, коли аналіз зв'язку між змінними X та Y, за даними кореляційної таблиці, приводить до вибору форми кореляційної залежності Y на X у вигляді показникової функції
   $$\bar{y}_x = ba^x, \tag{4}$$
   а при розгляді регресії X на Y –показникової функції
   $$\bar{x}_y = dc^y. \tag{5}$$

4. **Коренева кореляція.**
   Припустимо, що аналіз залежності між змінними X і Y , вираженої кореляційною таблицею, приводить до вибору форми кореляційної залежностіY на X у вигляді рівняння
   $$\bar{y}_x = a\sqrt{x} + b, \tag{6}$$
   а у випадку регресії X на Y – рівняння
   $$\bar{x}_y = c\sqrt{y} + d. \tag{7}$$

За побудованою кривою регресії $\bar{y}_x = f(x)$ (або $\bar{x}_y = (y)$) можна оцінити відхилення значень випадкової величини Y від кривої регресії $\bar{y}_x$ (або значень випадкової величини X від кривої регресії $\bar{x}_y$). Зокрема, обчислюють дисперсію величини Y відносно кривої регресії Y на X:

$$\sigma^2(y, \bar{y}_x) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l}(y_j - f(x_i))^2 n_{ij} = \frac{\Delta}{n}, \tag{8}$$

За міру розсіяння значень випадкової величини Y від кривої регресії $\bar{y}_x$ можна також взяти, наприклад, суму квадратів відхилень δ^2 умовних середніх

$$\overline{y_{xi}} = \frac{1}{n_i}\sum_{j=1} y_j n_{ij} \tag{9}$$

обчислених за даними кореляційної таблиці, від значень $f(x_i)$ функції регресії:

$$\delta^2 = \sum_{i=1}^{k}\delta_i^2 n_i = \sum_{i=1}^{k}|\overline{y_{xi}} - f(x_i)|^2 n_i. \tag{10}$$

3 Програмна реалізація.

Свою програму я реалізувала за допомогою мови програмування C#, використовуючи інтерфейс $WindowsForms$.

3.1 Інтерфейс користувача.

Запустивши програму, користувач може побачити вікно, на якому висвітлюється кореляційна таблиця частот, запропонована у варіанті 18. За бажання користувач може відредагувати будь-яку клітинку цієї таблиці. Знизу таблиці є кнопка "Обчислити функцію", що відповідає за побудову таблиць умовних середніх $\bar{y}_{xi}$ $(\overline{i = 1, k})$, та $\bar{x}_{yj}$ $(\overline{j = 1, l})$, а також за побудову поля кореляції, а саме точок $M_i(x_i; \bar{y}_{xi})$, $i = 1, ..., k$, та приблизного графіку регресії на координатній площині.

Після цього користувач може перевірити будь-яку з наступних гіпотез про форму кореляційного зв'язку, а саме гіпотези про гіперболічну, експоненційну, кореневу та параболічну кореляції. Після натискання відповідних кнопок для перевірки відповідних гіпотез, на екрані появляється панель, на якій висвітлюється система рівнянь відносно невідомих параметрів (a, b для гіперболічної, кореневої та експоненційної гіпотези або a, b, c для параболічної гіпотези), обчислені ці параметри, гіпотетична функція відносно знайдених параметрів, а також обчислені дисперсія величини Y відносно кривої регресії Y на X та сума квадратів відхилень δ^2 умовних середніх, від значень $f(x_i)$ функції регресії. Також після натискання цієї кнопки на панелі графіка появляється поле, при натисканні на яке, на координатній площині появляється графік знайденої функції.

3.2 BackEnd

Для отримання необхідних результатів була розроблена наступна система класів:

3.2.1 Клас $CorrelationTable$

- клас для представлення кореляційної таблиці у вигляді двовимірного масиву. У класі реалізовані наступні методи:

- $CountXMean$ – метод для знаходження умовних середніх $\bar{x}_{yj}$ $(\overline{j=1,l})$;
- $CountYMean$ – метод для знаходження умовних середніх $\bar{y}_{xi}$ $(\overline{i=1,k})$;
- $CountHyperbola$ – метод для знаходження коефіцієнтів системи рівнянь для гіпотези про вибір форми кореляційної залежності Y на X у вигляді рівняння гіперболи;
- $CountExp$– метод для знаходження коефіцієнтів системи рівнянь для гіпотези про вибір форми кореляційної залежності Y на X у вигляді показникової функції;
- $CountSqrt$ – метод для знаходження коефіцієнтів системи рівнянь для гіпотези про вибір форми кореляційної залежності Y на X у вигляді кореневої функції;
- $CountParab$ – метод для знаходження коефіцієнтів системи рівнянь для гіпотези про вибір форми кореляційної залежності Y на X у вигляді функції параболи;
- $CountSigma$ – метод для обчислення дисперсії величини Y відносно кривої регресії Y на X, метод приймає параметром делегат на функцію регресії;
- $CountDelta$ – метод для обчислення суми квадратів відхилень δ^2 умовних середніх обчислених за даними кореляційної таблиці, від значень $f(x_i)$ переданої параметром функції регресії;
- $print\_sigma$, $print\_delta$ – метод для виведення дисперсії та суми квадратів відхилень на екран;
- $CountAB$ – статична функція для обчислення коефійієнтів функції регресії.

3.2.2 Клас $Function$

- абстрактний клас для представлення функції регресії. Клас характеризується наступними полями та методами:

- Поле Res – список коефіцієнтів системи рівнять;
- Поле $func$ – делегат на функцію регресії;
- Поле ab – коефіцієнти функції регресії;
- Методи $PrintEquolity$, $PrintCoefs$, $PrintFunction$ – методи для виводу відповідної інформації на екран;

3.2.3 Класи $ExponentialFunction$, $HyperbolicFunction$, $ParabolicFunction$, $RootFunction$

- класи похідні від класу $Function$ реалізовані для представлення відповідних кореляцій.

3.2.4 Клас $Form$

Клас похідний від класу Form, що представляє вікно, що формолює інтерфейс користувача. За допомогою властивостей, доступных в класі Form, я визначаю зовнішній вигляд, розмір, колір и функції керування вікнами створеного вікна, а також взаємозвязок між панелями.

4 Отримані результати (графічні та числові) та їх аналіз.(Варіант 18)

Кореляційна таблиця

X|Y	1	2	4	6	9	11	12
3						7	31
4				2	21	4	
5			4	12	6		
7		3	22	5			
10	4	20					
12	23						

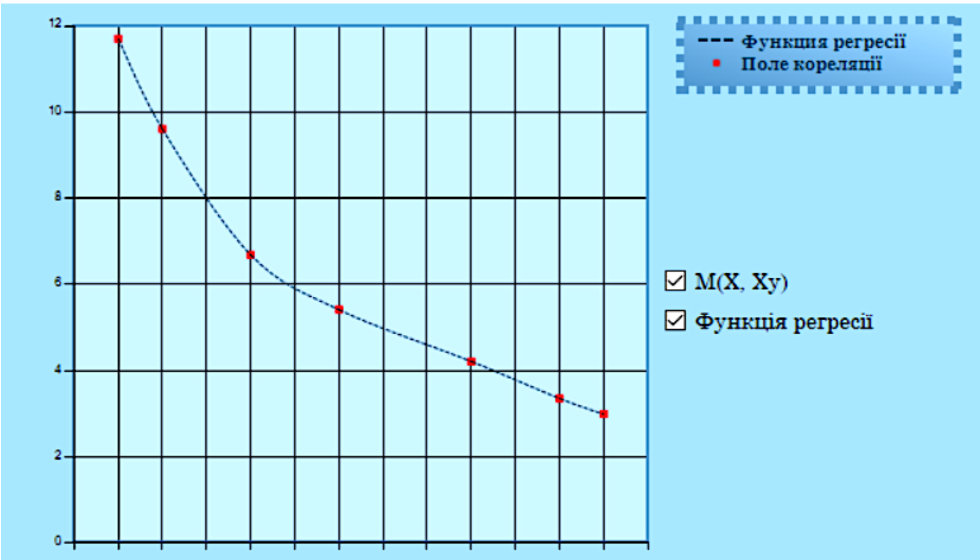
Умовних середніх $\bar{x}_{yj}$:

x	**1**	**2**	**4**	**6**	**9**	**11**	**12**
yx	**11.7**	**9.61**	**6.69**	**5.42**	**4.22**	**3.36**	**3.00**

Умовних середніх $\bar{y}_{xi}$:

y	**3**	**4**	**5**	**7**	**10**	**12**
xy	**11.8**	**9.07**	**6.45**	**4.13**	**1.83**	**1.00**

Поле кореляції:



4.1 Гіпотеза про гіперболічну кореляцію

Отримані результати:

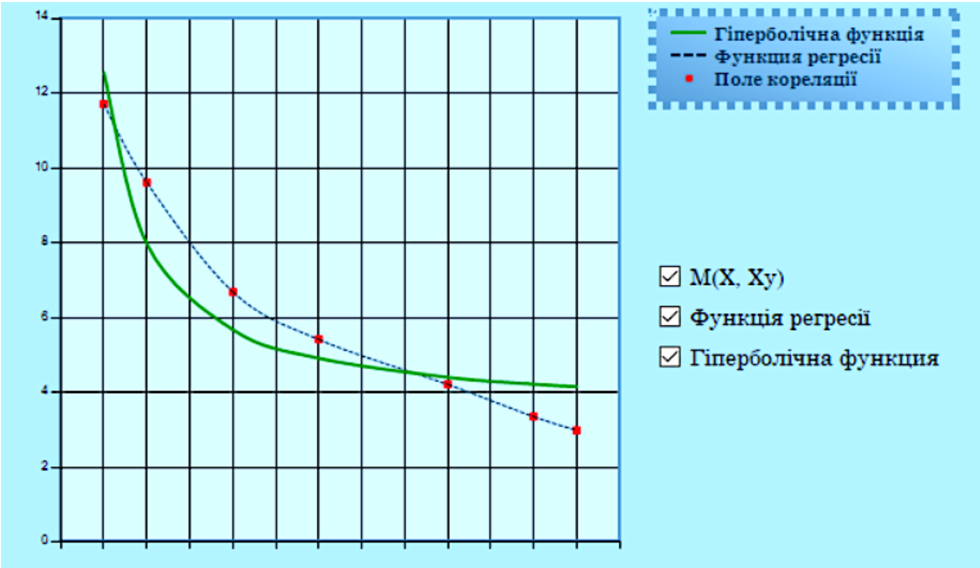
Гіперболічна функція

54.75 * a + 164.00 * b = 1058.00
35.54 * a + 54.75 * b = 510.95

a = 9.14; b = 3.40; y(x) = 9.14 / x + 3.40;	Sigma = 1.46; Delts ^ 2 = 162.85;

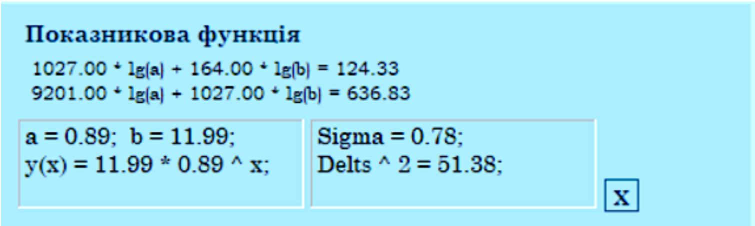
X

Графік гіперболічної кривої регресії:

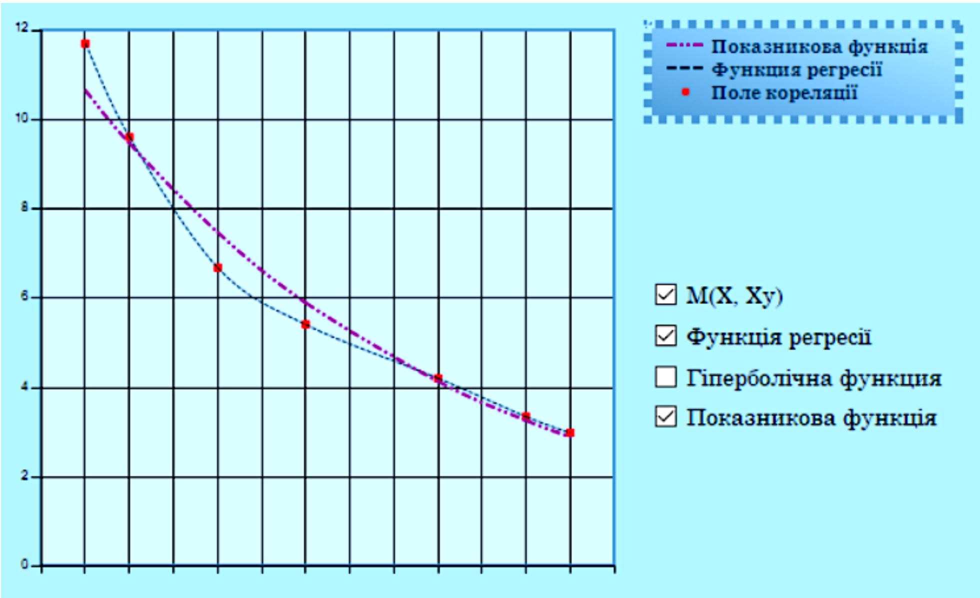


4.2 Гіпотеза про показникову кореляцію

Отримані результати:

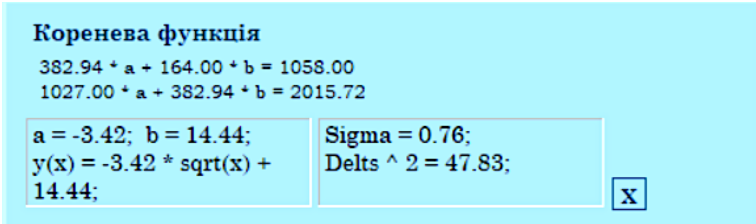


Графік показникової кривої регресії:

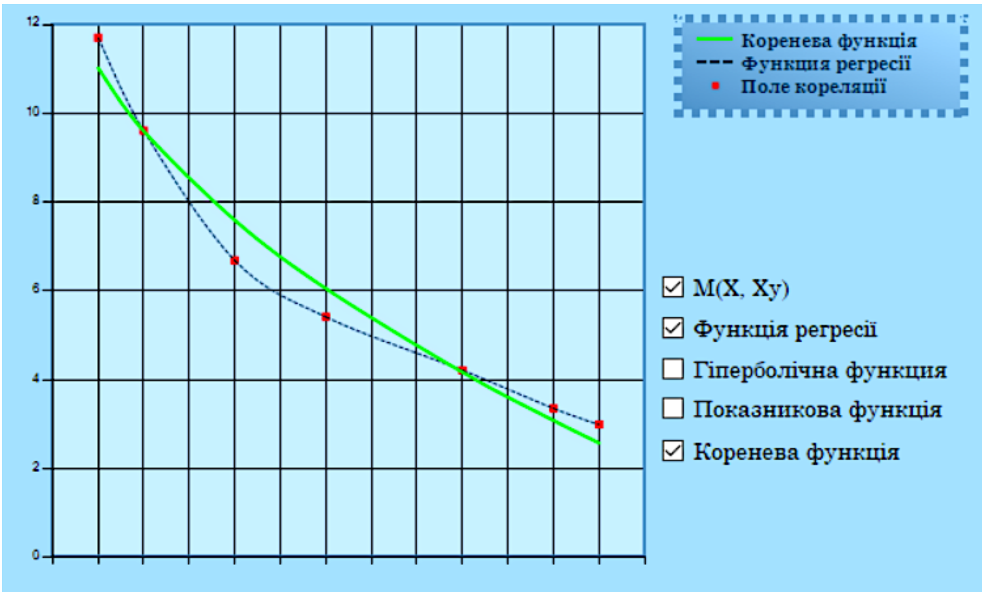


4.3 Гіпотеза про кореневу кореляцію

Отримані результати:

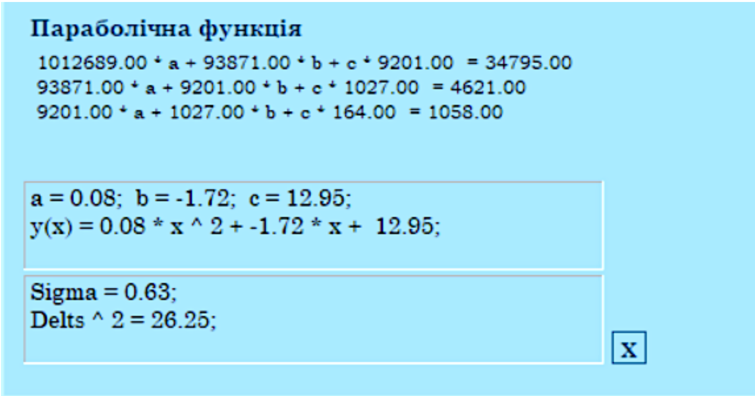


Графік кореневої кривої регресії:

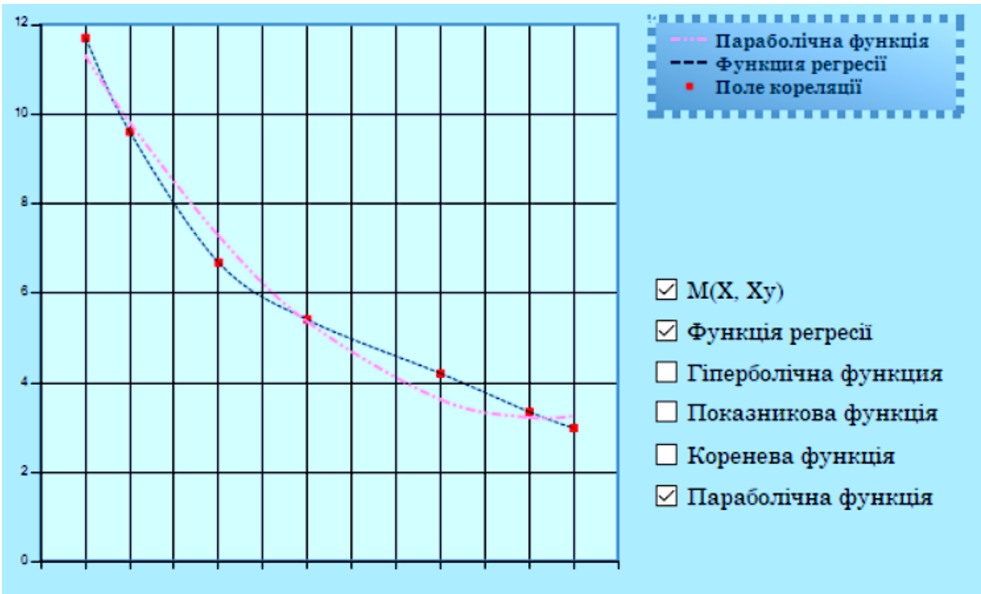


4.4 Гіпотеза про параболічну кореляцію

Отримані результати:



Графік параболічної кривої регресії:



4.5 Аналіз отриманих результатів

Проаналізувавши отримані результати, можна зробити висновок, що дані запропоновані у кореляційній таблиці варіанту 18 у порівнянні з прийнятими гіпотезами про вигляд функцій кореляції з найменшим відхиленням є **параболічною**.

5 Висновки.

Під час виконання цього індивідуального завдання я застосувала свої знання для перевірки гіпотез про вигляд кривої регресії та обчислення щільність отриманого кореляційного звязку. Я перевірла результати виконання моєї програми вручну та переконалася в правильності її обчислень.