Міністерство Освіти та Науки України
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА
Факультет прикладної математики та інформатики
Кафедра дискретного аналізу та інтелектуальних систем

ЗВІТ ПРО ВИКОНАННЯ
ІНДИВІДУАЛЬНОГО ЗАВДАННЯ №3

З дисципліни
Теорія Імовірності та Математична Статистика

Виконав студент
Групи ПМІ-24
Кутко Остап

2025

Постановка задачі:

1. За даними кореляційної таблиці обчислити умовні середні y_{xi} (i = 1,…,k).
2. Побудувати поле кореляції, тобто нанести точки Mi(xi; y_{xi}), i = 1,…,k , на координатну площину, та емпіричну лінію регресії.
3. Побудувати лінійне рівняння регресії та намалювати графік.
4. Обчислити коефіцієнт детермінації та перевірити адекватність побудованої лінійної моделі.
5. Обчислити вибірковий лінійний коефіцієнт кореляції.
6. За рівня значущості α перевірити значущість коефіцієнта кореляції.
7. Зробити припущення про вигляд функції нелінійної регресії (парабола, гіпербола і т.д.). В залежності від вигляду функції регресії скласти відповідну систему рівнянь. Розв'язати її і знайти невідомі параметри вибраної функції нелінійної регресії.
8. Записати рівняння кривої регресії Y на X : y = f(x) та побудувати її графік
9. Перевірити адекватність побудованої нелінійної моделі за F-критерієм
10. За моделлю з найменшою залишковою варіацією Q_o обчислити прогнозоване значення y* при заданому значенні x*

Короткі теоретичні відомості

Коваріацією між випадковими змінними ξ та η називають сподівання добутку відхилень цих змінних від своїх сподівань (або другому змішаному центральному моменту цих змінних), тобто

$$cov(\xi,\eta) = E\left[(\xi - E\xi)(\eta - E\eta)\right].$$

Кореляцією між випадковими змінними ξ та η називають відношення коваріації між змінними ξ, η до стандартів цих змінних $\rho(\xi,\eta) = \dfrac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}}$.

Якщо ξ та η незалежні випадкові змінні, то $\rho(\xi,\eta) = 0$.

Регресією випадкової змінної η на ξ називають добуток кореляції між цими змінними на відношення стандарту випадкової змінної η до стандарту випадкової змінної ξ

$$R(\eta/\xi) = \rho(\xi,\eta)\frac{\sqrt{D\eta}}{\sqrt{D\xi}}.$$

Для визначення вигляду функції регресії будують точки $(\bar{\bar{x}}_i, y_{xi})$ та з'єднують їх ламаною, яка називається **емпіричною лінією регресії.**

Компоненти випадкового вектора (ξ,η) лінійно-залежні, якщо вони пов'язані співвідношеннями $\eta = a\xi + b$ або $\xi = c\eta + d$

Компоненти випадкового вектора (ξ,η) називають некорельованими, якщо кореляція між ними дорівнює нулю.
Різниця значень y_i* і y_{xi} називається помилкою регресійної моделі і позначається e_i.

Вибіркова коваріація між координатами вектора (ξ,η) визначається так:
$$c_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
Вибіркова кореляція –
$$r_{12} = \frac{c_{12}}{s_1 s_2} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{i=1}^{k} n_i (x_i - \bar{x})^2}\sqrt{\sum_{j=1}^{l} m_j (y_j - \bar{y})^2}}$$
вибіркова регресія другої координати на першу
$$b_{21} = r_{12}\frac{s_2}{s_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$
вибіркова регресія першої координати на другу
$$b_{12} = r_{12}\frac{s_1}{s_2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$
Вибірковим аналогом дисперсії випадкової змінної
$\zeta = -a\xi + \eta, \quad (a > 0, a = const)$
є варіанса:
$$s_\zeta^2 = a^2 s_1^2 - 2\,a\,c_{12}\,s_1\,s_2 + s_2^2$$
Пряма $y - \bar{y} = b_{21}(x - \bar{x})$ називається **прямою регресії другої компонента вектора** (ξ,η) **на першу**.
Пряма $x - \bar{x} = b_{12}(y - \bar{y})$ називається **прямою регресії першої компонента вектора** (ξ,η) **на другу**.

Загальна варіація $Q = \sum_{i=0}^{k} (\overline{y_{x_i}} - \bar{y})^2 \, n_i$

Варіація регресії $Q_p = \sum_{i=0}^{k} (y_i^* - \bar{y})^2 \, n_i$

Варіація залишків $Q_o = \sum_{i=0}^{k} (\overline{y_{x_i}} - y_i^*)^2 \, n_i$

Коефіцієнт детермінації $R^2 = 1 - \frac{Q_o}{Q} = \frac{Q_p}{Q}$

Нелінійна регресія

$y = ax^2 + bx + c$ (параболічна кореляція другого порядку)

$y = ax^3 + bx^2 + cx + d$ (поліноміальна кореляція третього порядку)

$$\begin{cases} (\sum_{i=1}^{k} n_i x_i^4)a + (\sum_{i=1}^{k} n_i x_i^3)b + (\sum_{i=1}^{k} n_i x_i^2)c = \sum_{i=1}^{k} n_i \overline{y}_{x_i} x_i^2; \\ (\sum_{i=1}^{k} n_i x_i^3)a + (\sum_{i=1}^{k} n_i x_i^2)b + (\sum_{i=1}^{k} n_i x_i)c = \sum_{i=1}^{k} n_i \overline{y}_{x_i} x_i; \\ (\sum_{i=1}^{k} n_i x_i^2)a + (\sum_{i=1}^{k} n_i x_i)b + nc = \sum_{i=1}^{k} n_i \overline{y}_{x_i}. \end{cases}$$

$y = \frac{a}{x} + b$ (гіперболічна кореляція);

$$\begin{cases} a \sum_{i=1}^{k} \frac{1}{x_i} n_i + bn = \sum_{i=1}^{k} \overline{y}_{x_i} n_i; \\ a \sum_{i=1}^{k} \frac{1}{x_i^2} n_i + b \sum_{i=1}^{k} \frac{1}{x_i} n_i = \sum_{i=1}^{k} \frac{1}{x_i} \overline{y}_{x_i} n_i. \end{cases}$$

$y = a\sqrt{x} + b$ (коренева кореляція)

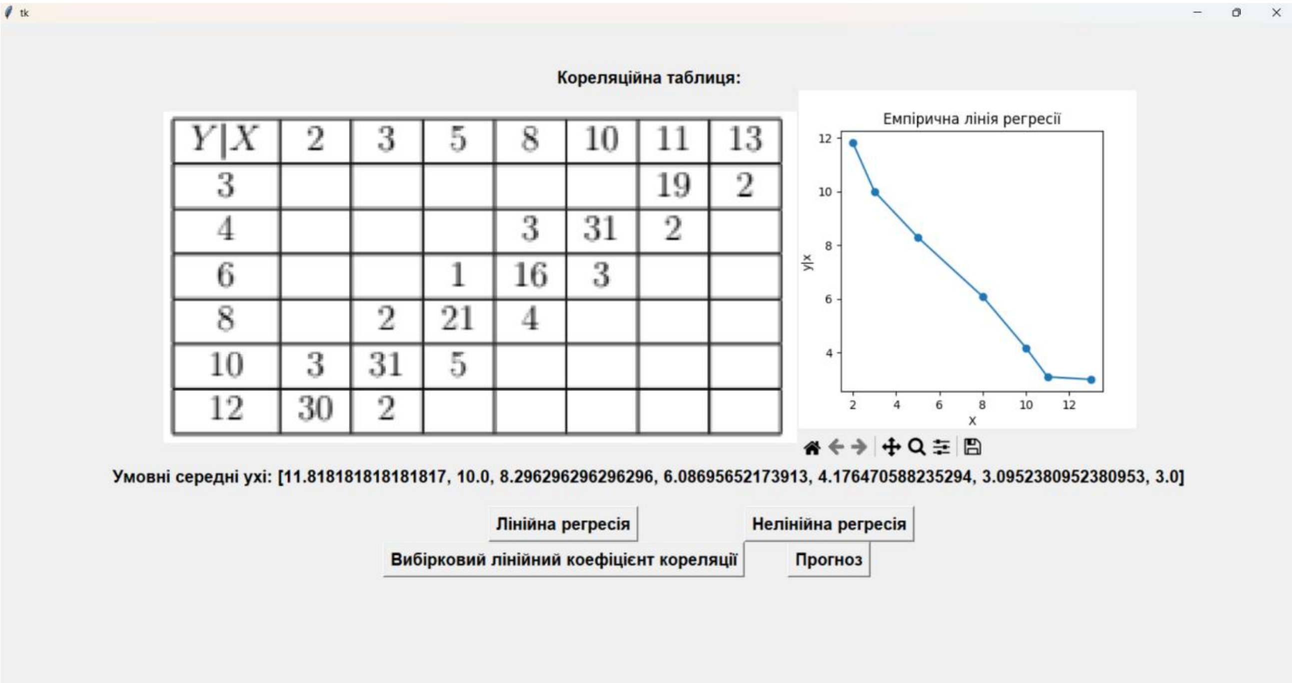
$$\begin{cases} a \sum_{i=1}^{k} n_i \sqrt{x_i} + bn = \sum_{i=1}^{k} \overline{y}_{x_i} n_i; \\ a \sum_{i=1}^{k} n_i x_i + b \sum_{i=1}^{k} n_i \sqrt{x_i} = \sum_{i=1}^{k} n_i \overline{y}_{x_i} \sqrt{x_i}. \end{cases}$$

$y = ba^x$ (показникова кореляція).

$$\begin{cases} \lg a \sum_{i=1}^{k} n_i x_i + n \lg b = \sum_{i=1}^{k} n_i \lg \overline{y}_{x_i}; \\ \lg a \sum_{i=1}^{k} n_i x_i^2 + \lg b \sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} n_i x_i \lg \overline{y}_{x_i}. \end{cases}$$

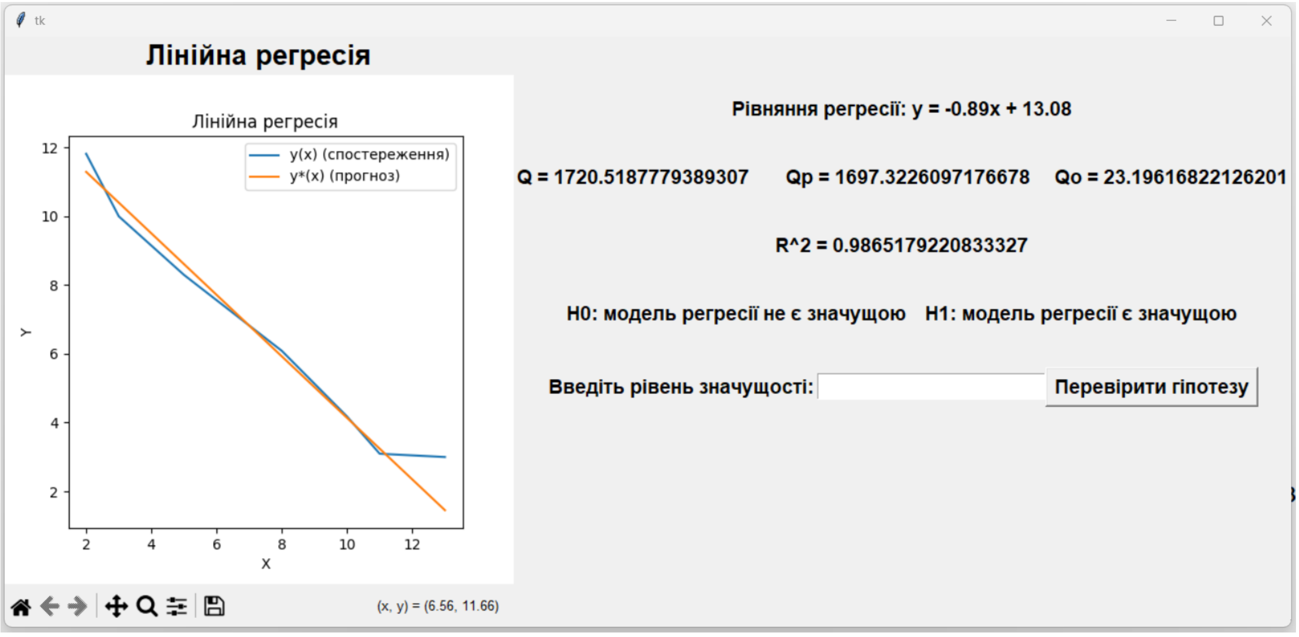
Програмна реалізація:

Запустивши програму, користувач бачить вікно, в якому є кореляційна таблиця відповідно до варіанту (6), а також одразу обчислені умовні середні y_{xi}, а також виведено емпіричну лінію регресії.

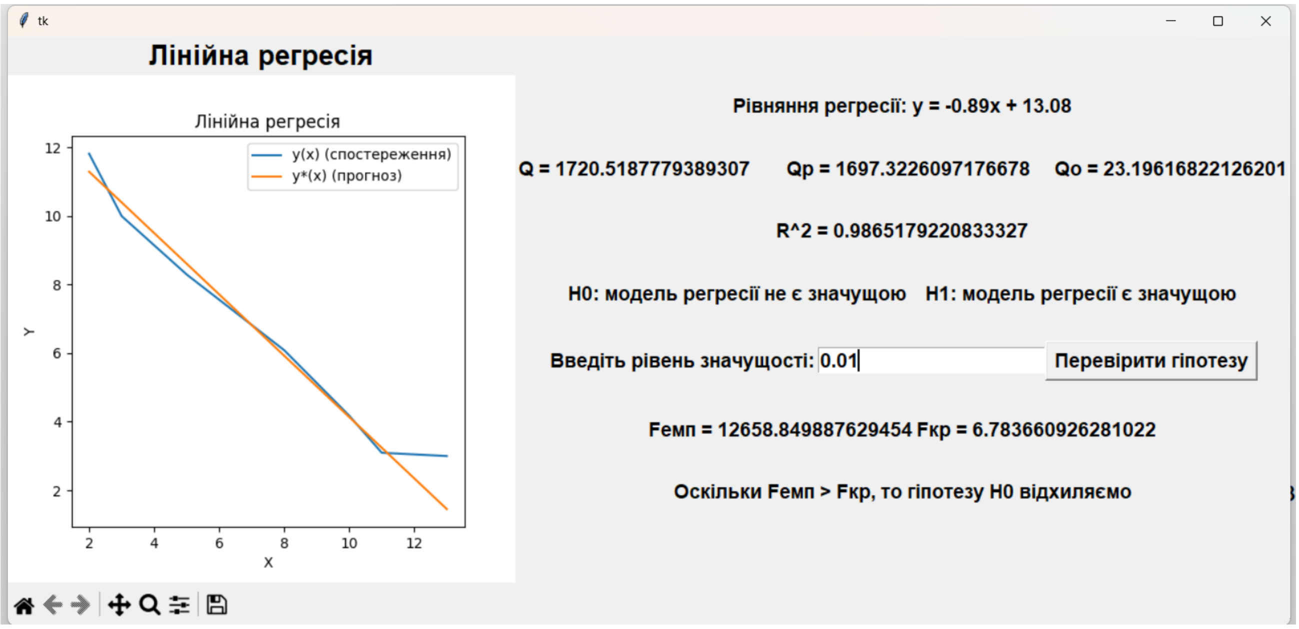


$Y \mid X$	2	3	5	8	10	11	13
3						19	2
4				3	31	2	
6			1	16	3		
8		2	21	4			
10	3	31	5				
12	30	2					

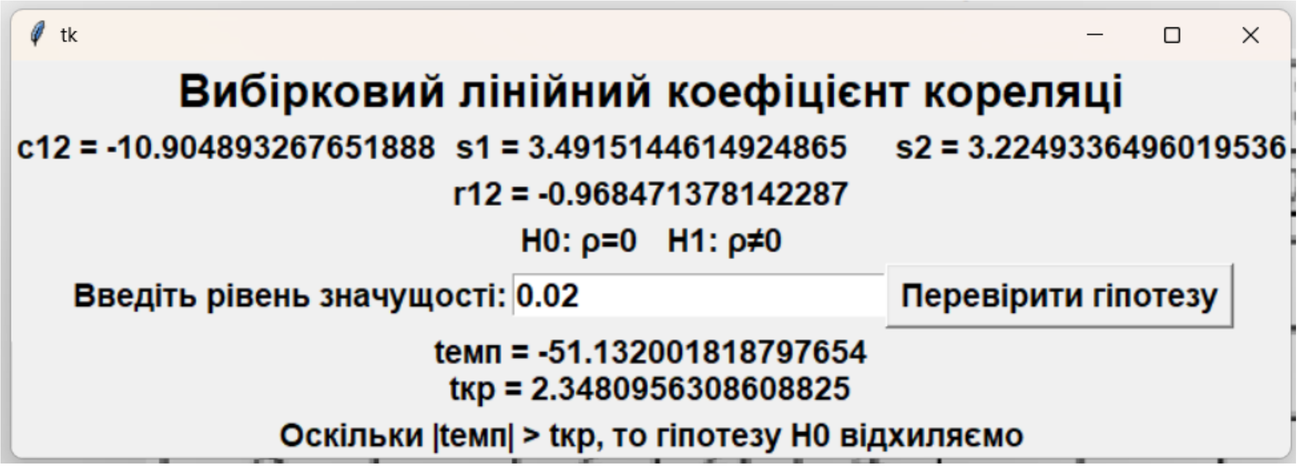
При натисканні на кнопку “Лінійна регресія”, програма обчислює значення параметрів рівняння a і b, а також загальну варіацію Q, варіацію регресії Qр і варіацію залишків Qо, коефіцієнт детермінації. Будує графік лінійної регресії. Після цього відкриває нове вікно з обрахованими значеннями, графіками і сформованими гіпотезами для перевірки статистичної значущості рівняння регресії.



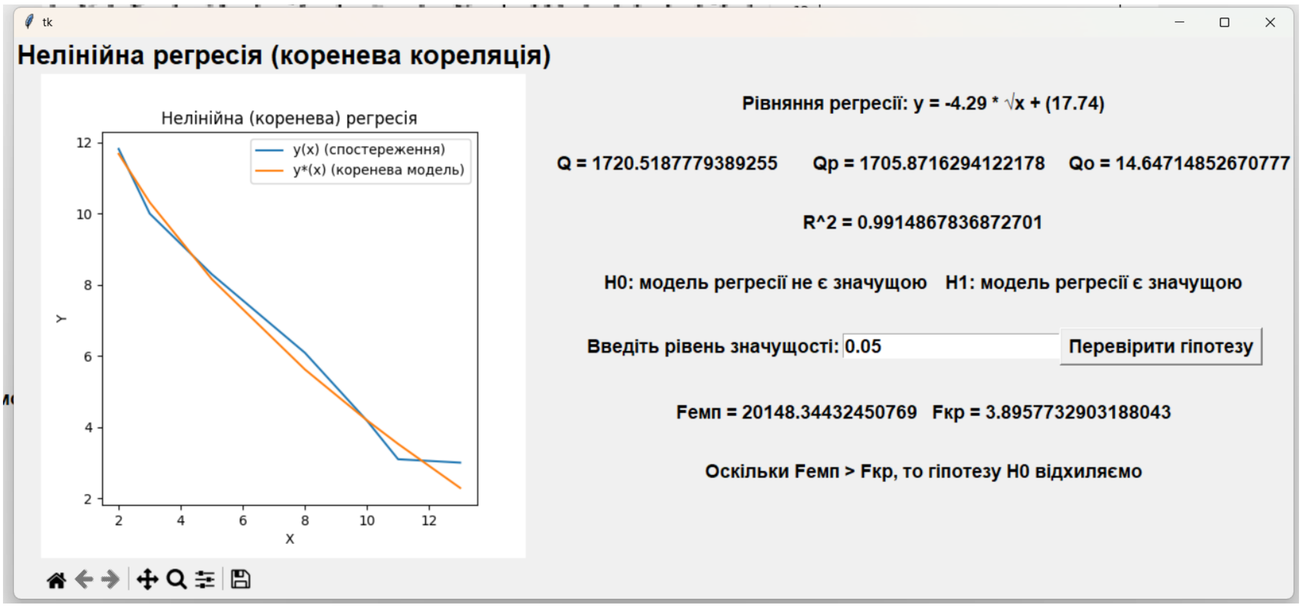
Користувач має ввести значення рівня значущості і тоді після натискання кнопки "Перевірити гіпотезу" за цим значенням програма обчислює F критичне, обчислює F емпіричне, порівнює їх і робить висновок про правильність гіпотези H0.



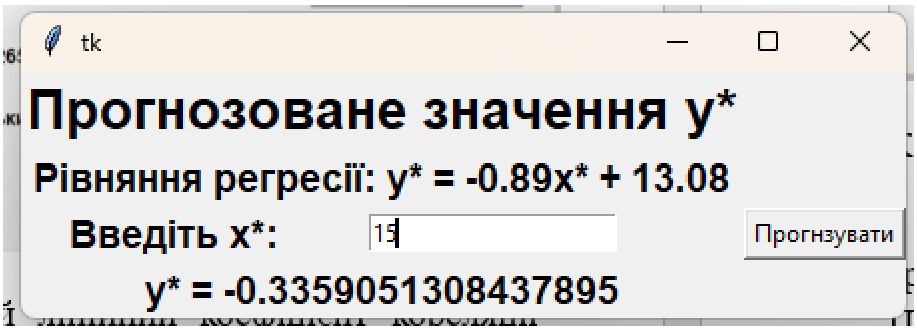
Після натискання кнопки "Вибірковий лінійний коефіцієнт кореляції" програма обраховує c12, s1, s2 і r12, формує гіпотезу і очікує на значення рівня значущості від користувача. Після того, як користувач ввів рівень значущості і натиснув на кнопку "Перевірити гіпотезу" обчислюються t емпіричне і t критичне з заданою значущістю, перевіряється нульова гіпотеза і виводиться висновок.



Кнопка “Нелінійна регресія”: пошук параметрів a і b, обчислення Q, Qp, Qo, R^2, F емпіричне, F критичне (після вказання рівня значущості), перевірка гіпотези, побудова графіка.
За виглядом графіка я використав кореневу кореляцію.



Кнопка “Прогноз”: прогнозує значення функції за кореневою регресією (бо в ній Qo менше ніж в лінійній).



Програму реалізовано на мові python, використано бібліотеки math, numpy, tkinter, matplotlib, scipy.

Висновок: У ході лабораторної роботи побудовано лінійну та нелінійну регресійні моделі, обчислено основні статистичні характеристики (коефіцієнти регресії, кореляції, варіації, R^2, F, t). Перевірено значущість моделей при заданому рівні значущості. Найбільш адекватною виявилася коренева модель, за якою виконано прогноз значення y∗ при заданому x∗.