



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Факультет прикладної математики  
Кафедра програмного забезпечення комп'ютерних систем

**ЗВІТ**  
**з лабораторної роботи № 3**  
**«Кореляційний аналіз, лінії регресії»**

**Дисципліна**  
**«Алгоритмічне забезпечення мультимедійних та інформаційно-  
пошукових систем»**

**Виконав:**

студент 2-го курсу, групи КП-12,  
спеціальності 121 – Інженерія  
програмного забезпечення  
*Якубишин Анатолій Сергійович*

**Перевірила:**

*Сушук-Слюсаренко Вікторія Ігорівна*

Київ – 2023

## Мета роботи:

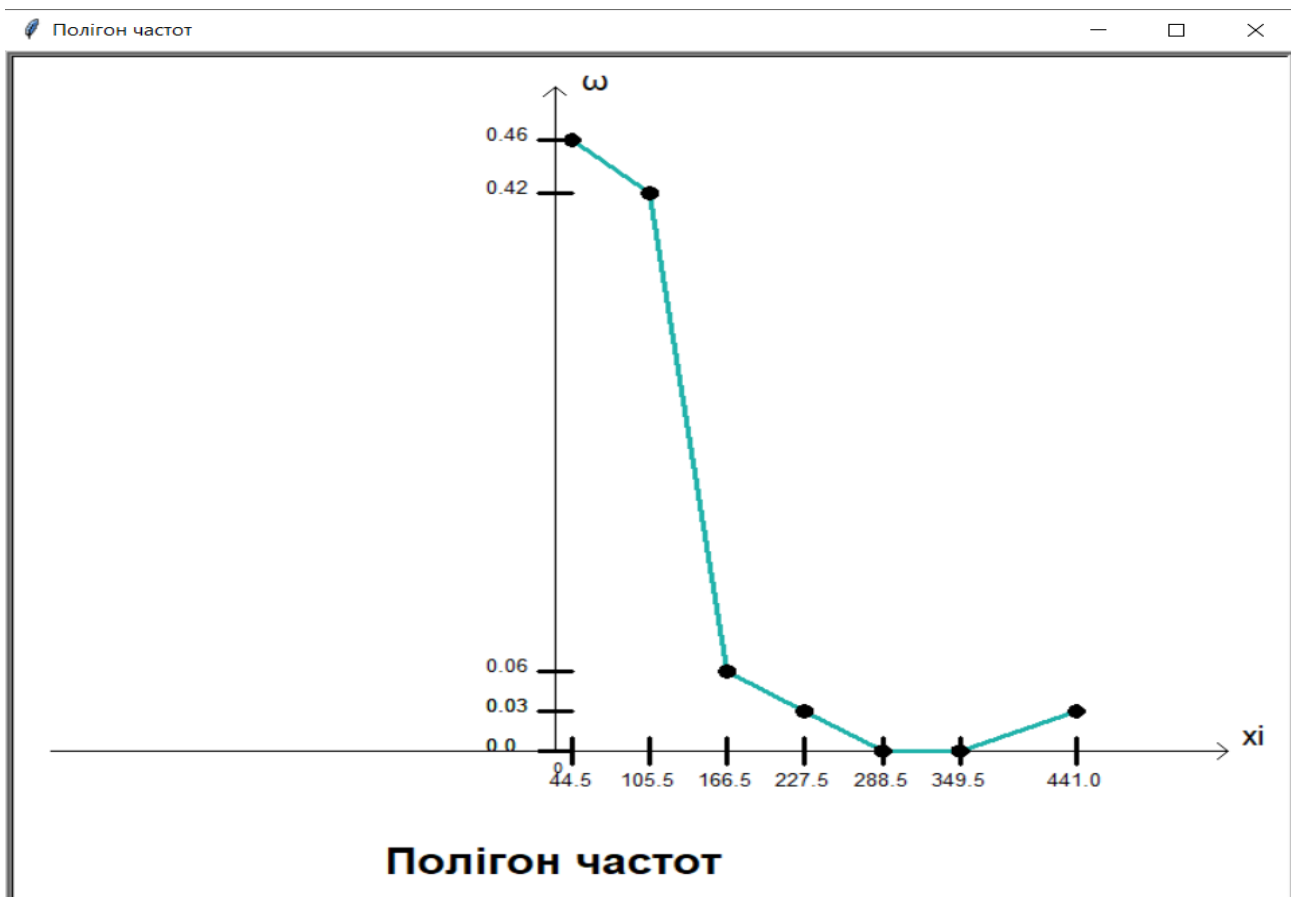
Навчитися досліджувати системи двох випадкових величин: будувати лінії регресії, обчислювати коефіцієнт кореляції, у разі нелінійної залежності, знаходити кореляційні відношення та індекс кореляції.

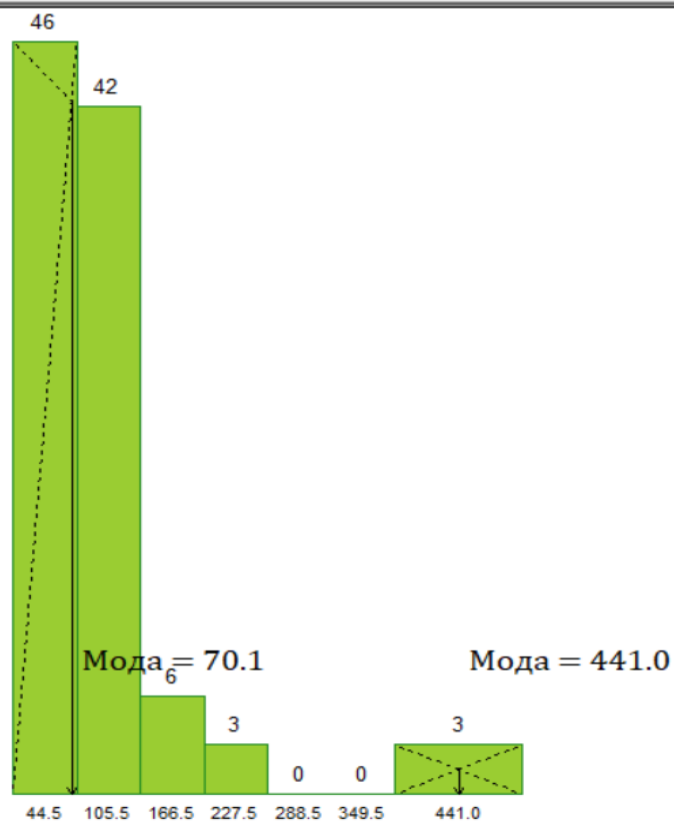
## Хід Роботи

За другу вибірку, тобто пов'язану величину візьму **відношення рівня інфляції в ЄС до рівня інфляції в США** кожен день від **2020-03-19** до **2022-14-12** (1000 днів) відповідно до періоду вимірюваного курсу.

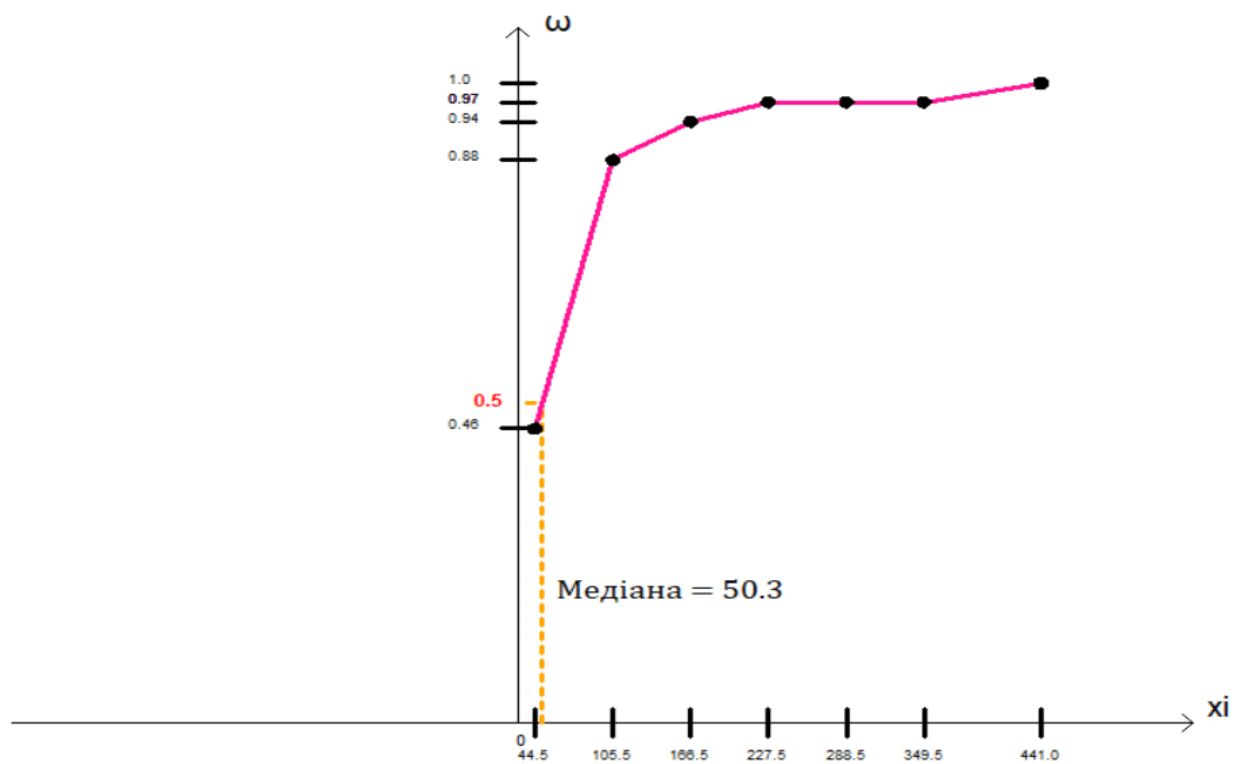
1) Використовуючи ПЗ для лаб 1 дістанемо такі дані:

```
Розмах вибірки 485.71428571
Ширина інтервалу: 60.71428571375
Візьмемо k = 61
x початкове = 14
[(44.5, 46), (105.5, 42), (166.5, 6), (227.5, 3), (288.5, 0), (349.5, 0), (441.0, 3)]
*****
Згрупований ранжований ряд зведемо в таблицю
| i | Кінці інтервалів | Середини інтервалів (xi) | Частота ni | Частість | Накопичена частота pi_нак | Накопичена частість |
|---|---|---|---|---|---|---|
1 | (14)-(75) | 44.5 | 46 | 0.46 | 46 | 0.46 |
2 | (75)-(136) | 105.5 | 42 | 0.42 | 88 | 0.88 |
3 | (136)-(197) | 166.5 | 6 | 0.06 | 94 | 0.94 |
4 | (197)-(258) | 227.5 | 3 | 0.03 | 97 | 0.97 |
5 | (258)-(319) | 288.5 | 0 | 0.0 | 97 | 0.97 |
6 | (319)-(380) | 349.5 | 0 | 0.0 | 97 | 0.97 |
7 | (380.0)-(502.0) | 441.0 | 3 | 0.03 | 100 | 1.0 |
Середнє = 94.8
Дисперсія 5644.2925
Середнє квадратичне відхилення 75.1285065737367
Коефіцієнт варіації 79.24947950816107 %
```





Гістограма



Кумулята

## Використовуючи ПЗ для лаб2 дістанемо такі дані:

```
Лабораторна 2
Невміщена точкова оцінка мат сподівання 94.8
Невміщена точкова оцінка дисперсії 5701.31

Підберу параметри квадратичної функціональної залежності y=ax^2+bx+c
a = 6e-06
b = -0.0042
c = 0.6732

y = 6e-06 * x^2 -0.0042 * x + 0.6732

Довірча ймовірність = 0.95

Довірчий інтервал для мат сподівання: 80.00062416991851 < a < 109.59937583008148
Довірчий інтервал для генеральної дисперсії: 4439.512314212602 < disp < 7771.5731354203745

Нульова гіпотеза Ho - генеральна сукупність має поліміальний закон розподілу з параметрами a,b,c що вказані вище, альтернативна - не узгоджується
Перевірю гіпотезу за критерієм Пірсона з рівнем значущості 0,05
|i|Кінці інтервалів| n1 | p1 | np1 | (ni-npi)^2 | ((ni-npi)^2)/(np1)|
1 14 - 75 46 0.5 50.0 16.0 0.32
2 75 - 136 42 0.4 40.0 4.0 0.1
3 136 - 197 6 0.07 7.0 1.0 0.14
4 197 - 502.0 6 0.08 8.0 4.0 0.5
Сума 100 1.05 105.0 - Kci спост 1.0628571428571432
Оскільки кількість рядків не більша кількості параметрів приймаю на 2, то беру k = 1
Kci критичне = 3.841458820694124
Узгоджується

Критична область: [ 3.841458820694124 ; +∞)

Висуну гіпотезу Ho - що генеральне середнє, дорівнює 99 , H1 - a < 99 Лівостороння критична область
t спостережене: = -0.007329801810194159
t критичне : = -1.6603911559963902
Критична область: (-∞; -1.6603911559963902 )
Прийнято нульової гіпотези нема
Потужність критерію: 0.54915918872909

При заданих рівнях значущості 0,05 та потужності 0,1 знайду мінімальний обсяг вибірки
n = 118
```

## 2) Об'єднаємо отримані дані в кореляційну таблицю

Співвідношення євро до долара у % (X)	Середини інтервалів  $y_i$  $x_i$	Співвідношення показника інфляції ЄС/США у відсотках				Всього n <sub>i</sub>	$\bar{y}_i$ (Групове середнє)
		14 - 75	75 - 136	136 - 197	197 - 502.0		
		44,5	105,5	166,5	349,5		
79-104	91,5	1	2	3	4	10	215,3
104 - 109	106,5	3	5	3	2	13	143,04
109 - 114	111,5	13	9	-	-	22	69,45
114 - 119	116,5	13	18	-	-	31	79,91
119 - 124	121,5	16	8	-	-	24	64,83
Всього n <sub>i</sub>		46	42	6	6	100	-
$\bar{x}_i$ (Групове середнє)		115,6	114	99	96,5	-	-

## 3) Знайдемо рівняння регресії X по Y, тобто залежність співвідношення євро до долара у відсотках до відношення рівня інфляції ЄС до США у відсотках.

Оберненої залежності Y по X не існує, бо залежність інфляції від курсу не є логічною, тому що інфляція залежить від переповнення каналів грошового обігу.

Отже, рівняння регресії Y по X шукати немає сенсу.

Обчислимо всі необхідні суми:

$$\sum_{i=1}^5 x_i \cdot n_i = 11280$$

$$\sum_{i=1}^5 x_i^2 \cdot n_i = 1279715$$

$$\sum_{j=1}^4 y_j \cdot n_j = 9574$$

$$\sum_{j=1}^4 y_j^2 \cdot n_j = 1457797$$

$$\sum_{i=1}^5 \sum_{j=1}^4 x_i \cdot y_j \cdot n_{ij} = 1043091$$

Знаходимо вибіркові характеристики і параметри рівняння регресії:

$$\bar{x} = \frac{11280}{100} = 112,80$$

$$\bar{y} = \frac{9574}{100} = 95,74$$

$$\mu = \frac{1043091}{100} - 95,74 \cdot 112,80 = -368,562$$

$$s_x^2 = \frac{1279715}{100} - 112,80^2 = 55,31$$

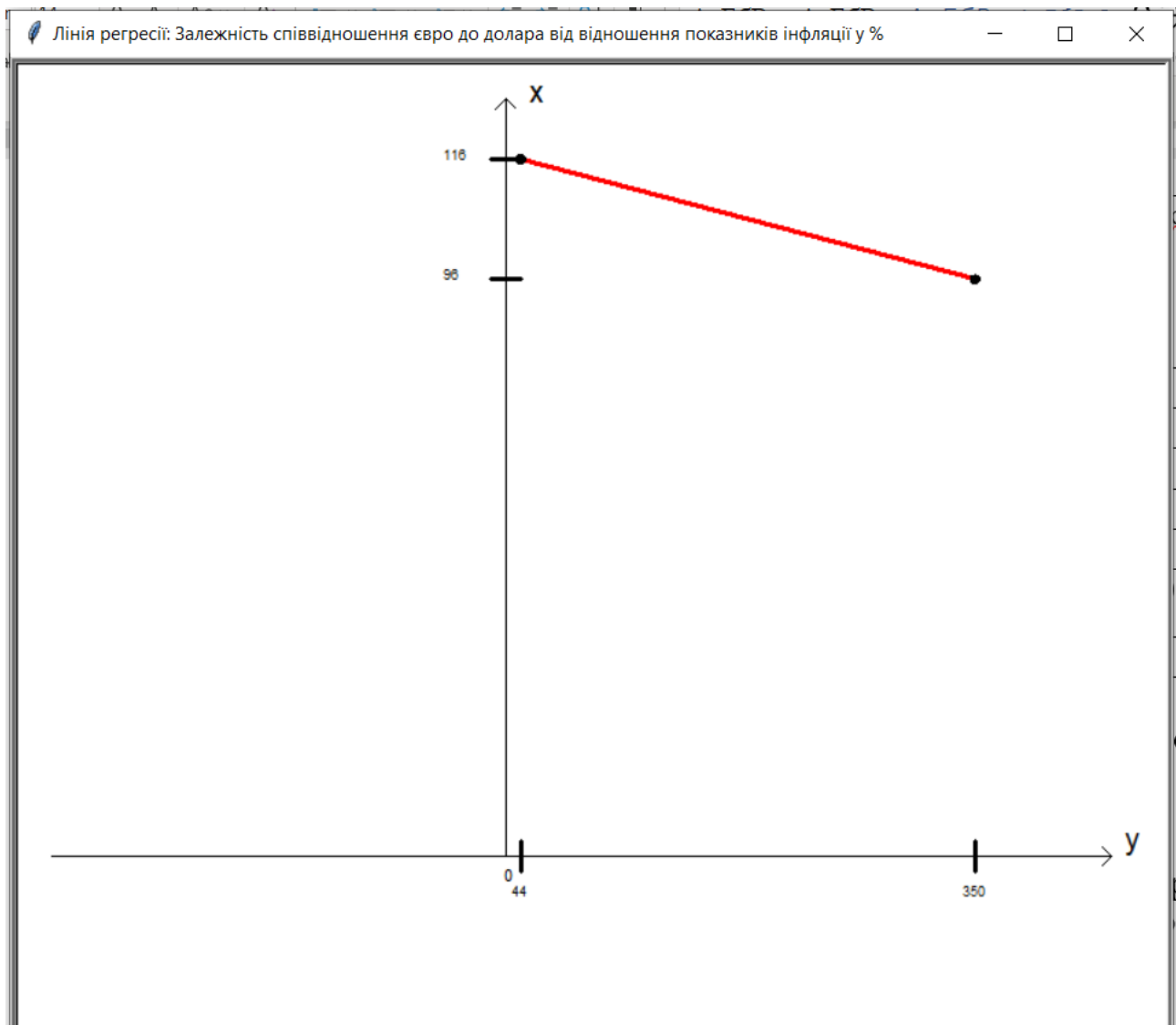
$$s_y^2 = \frac{1457797}{100} - 95,74^2 = 5411,82$$

$$b_{xy} = \frac{-368,562}{5411,82} = -0,068$$

$$x_y - \bar{x} = b_{xy}(y - \bar{y})$$

$$x_y - 112,80 = -0,068(y - 95,74)$$

$$x_y = 119,31 - 0,068y$$



$$b_{yx} = \frac{-368,562}{55,31} = -6,66357$$

$$r = -\sqrt{6,66357 \cdot 0,068} = -0,673144$$

Отже, зв'язок між змінними обернений і достатньо тісний.

Вивід програми:

```
Лабораторна 3  
Рівняння регресії X по Y:  
x = 119,31-0,068y  
вибірковий коефіцієнт кореляції = -0.673144
```

Перевіримо значущість коефіцієнта кореляції на рівні  $\alpha = 0,05$

$$\text{Обчислю статистику критерію } t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,673144\sqrt{98}}{\sqrt{1-0,673144^2}} = 9,01$$

$$\text{Знайду критичне значення статистики } t_{0,95;98} = 1,984$$

Оскільки  $t > t_{0,95;98}$ , то коефіцієнт кореляції між величинами  $X$  та  $Y$  значно відмінний від 0. А отже, велечини пов'язані між собою.

Побудуємо довірчий інтервал для генерального коефіцієнта кореляції та генерального коефіцієнта регресії з надійністю 0,95:

Для знаходження інтервалу для генерального коефіцієнта кореляції використаємо  $z$  – перетворення Фішера:

$$z = \frac{1}{2} \ln\left(\frac{1-0,673144}{1+0,673144}\right) = -0,81647$$

$$t_{0,05} = 1,96$$

$$-0,81647 - 1,96 \frac{1}{\sqrt{98}} \leq M(x) \leq -0,81647 + 1,96 \frac{1}{\sqrt{98}}$$

$$-1,01446 \leq M(x) \leq -0,61848$$

$$\frac{e^{-1,01446} - e^{1,01446}}{e^{-1,01446} + e^{1,01446}} \leq \rho \leq \frac{e^{-0,61848} - e^{0,61848}}{e^{-0,61848} + e^{0,61848}}$$

$\Rightarrow$

$$-0,7473 \leq \rho \leq -0,55$$

$$-0,068 - 1,984 \frac{55,31\sqrt{1-0,068^2}}{5411,82\sqrt{98}} \leq \beta_{xy}$$

$$\leq -0,068 - 1,984 \frac{55,31\sqrt{1-0,068^2}}{5411,82\sqrt{98}}$$

$$-0,07 \leq \beta_{xy} \leq -0,06595$$

Оскільки залежність лінійна, то індекс кореляції та кореляційне відношення не знаходжу.

### **Висновки**

У ході лабораторної роботи були розглянуті дві статистичні величини співвідношення курсу євро до долара у відсотках та співвідношення показника інфляції у відсотках. Дослідження показує, що величини тісно пов'язані між собою. Було знайдено лінію регресії  $X$  по  $Y$  та довірчі інтервали для параметрів зв'язку.